AD_____

Award Number: DAMD17-99-1-9150

TITLE: Understanding Single-Stranded Telomere End Binding by an Essential Protein

PRINCIPAL INVESTIGATOR: Emily M. Anderson, Ph.D.
Deborah S. Wuttke, Ph.D.

CONTRACTING ORGANIZATION: University of Colorado
Boulder, Colorado 80309

REPORT DATE: August 2002

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2002	Final (1 Aug 99 - 31 Jul 02)

4. TITLE AND SUBTITLE
Understanding Single-Stranded Telomere End Binding by an Essential Protein

5. FUNDING NUMBERS
DAMD17-99-1-9150

6. AUTHOR(S)
Emily M. Anderson, Ph.D.
Deborah S. Wuttke, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Colorado
Boulder, Colorado 80309

E-MAIL:
Emily.Anderson@Colorado.EDU

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
report contains color

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Cdc13p is an essential protein from *S. cerevisiae* that binds to the single-stranded ends of telomeres with high specificity and affinity. Cdc13p perform functions in concert with two protein complexes - protecting the end of the chromosome from degradation and regulating telomere length through the enzyme telomerase. Cdc13p binds yeast single-stranded telomeric DNA (sstelo DNA) *in vitro* with high affinity (K_d=0.3 nM). The modular DNA-binding domain of the protein has been mapped by deletion analysis and proteolysis.

We are investigating the structural and biochemical basis for high affinity binding and sequence specificity of the single-stranded DNA binding domain. We have used heteronuclear, multidimensional NMR spectroscopy to solve the high-resolution solution structure of the domain bound to a single-stranded telomeric DNA 11-mer. We have used isotope filtered and select/filter NOE experiments to determine the single-stranded DNA conformation in the complex and observe protein-DNA contacts. Also, the thermodynamic contribution of amino acids at the interface has been probed by site-directed mutagenesis and filter-binding.

14. SUBJECT TERMS	15. NUMBER OF PAGES
breast cancer, telomere, protein, DNA binding	75
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

Telomeres are the nucleoprotein complexes that protect the ends of linear eukaryotic chromosomes. Telomere replication and length regulation are controlled by the enzyme telomerase and a suite of telomere binding proteins. Anomalous telomeric replication is implicated in most forms of human cancer. Telomere metabolism is thus an active field in basic research for the eventual goal of developing inhibitors or modulators of telomere replication for cancer therapy. Cdc13p is an essential protein from the budding yeast *Saccharomyces cerevisiae* whose role is to protect the end of the chromosome from degradation and to load telomerase in concert with the protein Est1p. Biochemically, Cdc13p binds to single-stranded yeast telomeric DNA with high affinity and specificity. We are investigating the structural basis for high affinity binding and sequence specificity of the DNA binding domain. One aspect of this research involves solving the high resolution solution structure of the domain complexed to DNA using heteronuclear, multidimensional NMR. Biochemical techniques are also being employed, including mapping regions of the domain in proximity to the DNA by photocrosslinking, investigating sequence specificity using libraries of DNA with varying sequences, and determining the thermodynamic contributions of amino acids that contact DNA.. The advantage of studying this protein using yeast as a model organism is the power of combining structure, biochemistry, and genetics all in one system.

BODY

All of the technical goals have been completed in the last year. Technical objective 1, outlined below, was completed in full as of the report submitted two years ago.

Technical Objective 1:

Express and purify DNA binding constructs	2 Months
Conduct binding assays with site-randomized DNA	4 Months
Conduct CD experiments of protein folding and DNA binding	1 Month

An optimized DNA-binding domain construct had been delineated using proteolysis and MALDI mass spectrometry. This construct had been subcloned, expressed and purified in high yield, suitable for high resolution structural characterization. The construct binds DNA with much higher affinity comparable to that reported for the full-length protein as measured by both gel-shift binding assays and nitrocellulose filter-binding assays. This work, as well as the work form the first subtask in technical objective 2, have been accepted for publication in *Nucleic Acids Research*. This manuscript is located in the report as Appendix 1.

Technical objective 2 is also complete, as outlined below.

Technical objective 2:

Conduct photocrosslinking/identify contacts	3 Months
Design mutants/test *in vitro* and *in vivo*	6 Months

Photocrosslinking experiments with the chromophore iodouracil substituted for thymine have been performed. As specific amino acids were not identified using this technique, mutagenesis to analyze the DNA-binding interface was designed in conjunction with the data collected in technical objective 3, and is described below.

Technical objective 3 involves determining the high resolution NMR solution structure of the domain and has also been completed. As stated in the annual progress report from two years ago, the focus of the research until now has been on the protein/DNA complex with the collaboration of another student in the laboratory, Rachel Mitton-Fry. This work has been published in the journal *Science* and is located in this report as Appendix 2. The original Technical Objective 3 is listed as follows:

Technical Objective 3:

Optimize solution conditions of sample for NMR spectroscopy	1 Month
Protein alone –	
Collect heteronuclear NMR data for resonance assignment	6 Months
Assign resonances in the protein domain	6 Months
Collect heteronuclear NMR data for distance restraints	1 Month
Determine family of structures that satisfy restraints	6-12 Months
Protein/DNA complex –	
Titrate DNA into protein and conduct NMR experiments	6-18 Months

In our preparations of the complex the DNA is unlabeled and is not observed in the isotope-selected experiments conducted to complete the structure of the protein. Last year I performed isotope-filtered NMR experiments to examine the unlabeled single-stranded DNA in the complex and it appears to be in a unique, extended conformation with 11 identifiable spin systems. To aid in assignment of the spin systems and determine the conformation of the single-stranded DNA, this year I have conducted experiments with various thymine bases substituted with uracil as well as site-specific ^{13}C-labeled samples. This work will be presented in a manuscript intended for *Nature Structural Biology*. I have also conducted isotope select-filter experiments to measure NOE contacts between the protein and DNA, which are shown in Figure 3 in the paper in Appendix 2. We mapped a DNA-binding interface or cleft on the protein structure which is consistent with other measurements on the complex such as chemical shift changes that occur upon binding, protection from hydrogen exchange, and a net positively-charged groove located on the surface of the protein. Mutations have been made based on this interface and we have tested the thermodynamic contributions of these interface residues

to binding. This work has been submitted to *PNAS* and the paper is Appendix 3 of this report.

. It should be noted that some of the subtasks in technical objective 3 have been completed in parallel by myself, while some have been completed by Rachel Mitton-Fry. In this respect completion of the entire project, with a total time frame of 5 years, has been accomplished within the scope of the granting period.

KEY RESEARCH ACCOMPLISHMENTS (THIS YEAR)

- Double-filtered isotope experiments with substitutions of thymine by uracil, along with site-specific ^{13}C labeling of nucleotides and ^{13}C HSQC experiments have allowed for assignment and calculation of the DNA conformation in the complex.
- Complete structures of the protein/DNA complex have been calculated.
- Point mutations have been made and tested for the thermodynamic effect on DNA recognition of residues at the DNA-binding interface.

REPORTABLE OUTCOMES

Abstracts: The work in progress has been presented as a poster at one meeting: IBC's Drug Discovery Technology 2002 Meeting in Boston, MA .

Presentations: This work has been presented as a talk at the University of Colorado RNA Club in May, 2002.

Manuscripts:

Anderson, E. M., Halsey, W. A., Wuttke, D. S. (2002) Delineation of the high-affinity single-stranded telomeric DNA-binding domain of *S. cerevisiae* Cdc13. *Nucleic Acids Res.*, in press.

Mitton-Fry, R. M., Anderson, E. M., Hughes, T. R., Lundblad, V., Wuttke, D.S. (2002) Conserved structure for single-stranded telomeric DNA recognition. *Science*, **296**, 145-147.

Anderson, E. M., Wuttke, D. S. (2002) Site-directed mutagenesis reveals the thermodynamic requirements for single-stranded DNA recognition by the telomere-binding protein Cdc13. *Proc. Natl. Acad. Sci. USA*, submitted.

Degrees awarded:

Emily Anderson
Ph.D., Biochemistry, University of Colorado, Boulder, CO, (to be awarded Oct. 3, 2002).
Thesis title: Recognition of Single-Stranded Telomeric DNA by Cdc13.
Advisor: Dr. Deborah S. Wuttke.

"Delineation of the High-Affinity Single-Stranded Telomeric DNA-Binding Domain of *S. cerevisiae* Cdc13"

Emily M. Anderson, Wayne A. Halsey, and *Deborah S. Wuttke

Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, CO, 80309-0215

* corresponding author

ABSTRACT

Cdc13 is an essential protein from *Saccharomyces cerevisiae* that caps telomeres by protecting the C-rich telomeric DNA strand from degradation and facilitates telomeric DNA replication by telomerase. *In vitro*, Cdc13 binds TG-rich single-stranded telomeric DNA with high affinity and specificity. A previously identified domain of Cdc13 encompassing amino acids 451-694 (the 451-694 DBD) retains the single-stranded DNA-binding properties of the full-length protein; however, this domain contains a large unfolded region identified in heteronuclear NMR experiments. Trypsin digestion and MALDI mass spectrometry were used to identify the minimal DNA-binding domain (the 497-694 DBD) necessary and sufficient for full DNA-binding activity. This domain was completely folded, and the N-terminal unfolded region removed was shown to be dispensable for function. Using affinity photocrosslinking to site-specifically modified telomeric single-stranded DNA, the 497-694 DBD was shown to contact the entire 11-mer required for high-affinity binding. Intriguingly, both domains bound single-stranded telomeric DNA with much greater affinity than the full-length protein. The full-length protein exhibited the same rate of dissociation as both domains, however, indicating that

the full-length protein contains a region that inhibits association with single-stranded telomeric DNA.

INTRODUCTION

Telomeres are nucleoprotein complexes that form the ends of eukaryotic chromosomes. They are composed of repetitive tracts of DNA and a suite of proteins that specifically recognize both the double-stranded region and the G-rich single-stranded 3' overhang of telomeric DNA (1). Telomeres perform various functions in the cell, including capping the end of the chromosome, protecting it from degradation and end-to-end fusion, and serving as a substrate for telomerase, the specialized reverse transcriptase that replicates telomeres (2).

Several strategies have been identified for telomere capping by telomere end-binding proteins (3). For example, in the hypotrichous ciliate *Oxytricha nova*, the heterodimeric telomere end-binding protein (TEBP) specifically recognizes and buries the single-stranded overhang (4,5). An end-binding protein with limited sequence similarity to *O. nova* TEBP has been identified both in the fission yeast *Schizosaccharomyces pombe* and in humans (6). Deletion of this protein (Pot1) in fission yeast results in loss of telomeric DNA, chromosomal missegregation, and reduced growth that could be bypassed by circularization of the chromosome. In mammalian cells, a TRF2-mediated duplex lariat structure at the terminus of the chromosome called a t-loop has been proposed to sequester the end of the chromosome (7).

Cdc13 is an essential telomere-capping protein from the budding yeast *Saccharomyces cerevisiae* that protects the C-rich telomeric strand from degradation (8-11). *In vivo*, the temperature sensitive mutant allele *cdc13-1* causes resection of the C-

rich strand at the non-permissive temperature, along with cell-cycle arrest and lethality (12). In independent genetic studies, *CDC13* has also been shown to be a positive and negative regulator of telomere length (13). Mutation of residue 252 of Cdc13 causes a failure in telomere replication, even though the catalytic function of telomerase is not impaired in such mutant strains (14-16). These mutant alleles of *CDC13* can be reciprocally suppressed by certain mutations of the *EST1* subunit of telomerase (16), suggesting that the positive regulatory role of Cdc13 *in vivo* is recruitment of the enzyme to telomeric chromatin. Consistent with these activities, Cdc13 is believed to be localized to the 3' single-stranded overhang at the telomere as it binds single-stranded yeast telomeric DNA with both high affinity and specificity (14,17). Cdc13 does not bind double-stranded telomeric DNA or single- or double-stranded DNA of random sequence, and it does not require a free 3' end for binding. Cdc13 has the same affinity for binding a free single-stranded 3' end, however, such that it could bind and localize to the very end of the 3' overhang *in vivo*. In fact, the DNA-binding function of Cdc13 in isolation has been shown to be active *in vivo*. Tethering of the telomerase components Est1p or Est3p to the telomere by fusion with the Cdc13 DNA-binding domain (DBD) restores immortality to senescing mutants of Cdc13 and results in longer telomeres, respectively (18,19). In addition, fusion of the DBD to the end-protection factor Stn1p restores cell viability in the absence of full-length, functional Cdc13, although these cells still undergo senescence since they are unable to recruit telomerase components (16). These experiments indicate that Cdc13 functions to localize key proteins to the telomere that are involved in telomere end protection and replication.

A DNA-binding domain of Cdc13 (451-694 DBD) was previously identified within the 924-residue full-length protein by deletion mapping and limited proteolysis, facilitating biochemical studies of Cdc13 bound to single-stranded telomeric DNA (20,21). The DNA-binding domain has no similarity to any sequence in the database. Notably, no sequence similarity can be detected between the 451-694 DBD and the other telomeric DNA end-binding proteins *S. pombe* and human Pot1 (6) or the heterodimeric *O. nova* telomere end-binding protein (TEBP) (22-24). Even in the complete absence of sequence similarity, the recent solution structure of the Cdc13 DBD characterized here revealed that this domain adopts the same fold as both TEBP and the predicted fold of Pot1 (25). This result suggests that these telomere-binding proteins are evolutionarily related and that structure-function studies of Cdc13 are directly relevant to telomere maintenance in other organisms.

Although the 451-694 DBD retains DNA-binding properties comparable to that of the full-length protein, as assessed by both biochemical (20,21) and genetic (16,18) studies, it does not represent the minimal, independently folded structural domain. To better understand structure-function relationships governing single-stranded DNA binding at telomeres, we have further characterized the biochemical and structural properties of the minimal DNA-binding domain.

MATERIALS AND METHODS

Production of recombinant DBDs

DNA encoding the Cdc13 single-stranded telomeric DNA-binding domains (amino acids 451-694 and 497-694) was PCR-amplified from a genomic clone (generously provided by the Lundblad lab, Baylor College of Medicine) and subcloned

into the T7 expression vector pET21a (Novagen) between the Nde1 and Xho1 restriction sites. Electrotransformed (1.8 V, 400 Ω, 25 μF) BL21(DE3) *E. coli* were grown in LB medium with 50 μg/L ampicillin at 37°C to an OD_{600} of 0.6 and induced with 1 mM IPTG at 22°C for 4-5 hours. Cells were harvested by centrifugation, resuspended in buffer A (50 mM potassium phosphate, pH 7.0, 50 mM NaCl, 0.5 mM Na_2EDTA, 0.02% NaN_3, and 2 mM DTT), and lysed by two passes through a French press (Aminco). Cell extract was cleared of DNA by precipitation with 0.1% polyethylenimine at 4°C for 30 minutes with stirring and then centrifugation. The addition of a PEI precipitation step in the purification protocol, which is included in the purification of every protein studied here, was necessary in order to remove endogenous *E. coli* nucleic acids that were non-specifically bound to the recombinant protein. The amount of non-specifically bound nucleic acid could be readily followed by monitoring the A260/A280 ratio in the UV/Vis absorption spectrum. Failure to remove the bound nucleic acids results in spurious binding features. The cleared supernatant was purified by ion exchange chromatography over a 5 mL HiTrap SP-Sepharose column (Pharmacia) by gradient elution with buffer B (buffer A with 1M NaCl). The protein eluted at approximately 60% B and is over 95% pure as estimated by Coomassie-stained SDS-PAGE. Yield of protein was typically 15-20 mg per liter cells.

Production of recombinant his-tagged 497-694 DBD

DNA encoding the 497-694 DBD was PCR-amplified and subcloned into the expression vector pET21a (Novagen), between the NdeI and XhoI restriction sites, in-frame with the C-terminal His-tag. BL21(DE3) *E. coli* were transformed, grown, and induced as described above. Cells were pelleted by centrifugation, resuspended in lysis

buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 10% glycerol, 0.5% Tween 20, 10 mM imidazole, 5 mM β-mercaptoethanol, 1 mM PMSF, 5 μM pepstatin A, 10 μM leupeptin, 100 μM antipain, and 200 μM chymostatin), and lysed by French press. Cellular debris was removed by centrifugation, and the supernatant was cleared of DNA by precipitation with 0.1% polyethylenimine at 4°C for 30-45 minutes with stirring, followed by centrifugation. The supernatant was purified by affinity chromatography chromatography using a 5 mL HiTrap Chelating HP column (Pharmacia) charged with nickel chloride. Equilibration/wash buffer is the same as lysis buffer but without protease inhibitors. Bound protein was eluted with a linear gradient of imidazole from 10 mM to 500 mM. Protein was estimated to be 95% pure as measured by Coomassie-stained SDS-PAGE.

Production of recombinant full-length Cdc13

Frozen SF9 cells infected with baculovirus expressing full-length His_6-Cdc13 (amino acids 1-924) were provided by the Lundblad laboratory. The protein was purified using nickel-NTA chromatography as described previously (14).

Limited trypsin cleavage of the 451-694 DBD

The 451-694 DBD (30 μM in 200 μL), alone or with 1 molar equivalent telomeric 11-mer (dGTGTGGGTGTG), was incubated with 0.4% w/w trypsin (Sigma) at room temperature in 50 mM potassium phosphate, pH 7.0, 350 mM NaCl, 0.25 mM Na_2EDTA, 0.02% NaN_3, and 2 mM DTT. At various time points, 8 μL aliquots of the reaction were withdrawn, diluted with 12 μL water and 5 μL SDS loading dye, and run on 15% SDS-PAGE visualized with Coomassie Brilliant Blue.

MALDI mass spectrometry of the limited trypsin cleavage products

For MALDI mass spectrometric analysis, the 451-694 DBD (30 μM in 200 μL) was incubated with 0.4% w/w trypsin in 10 mM Tris, 2 mM DTT at room temperature. Aliquots were withdrawn as described above for SDS-PAGE analysis. At the 10 minute time point, 2 μL of the reaction mixture was mixed with 2 μL of CHCA (α-cyano-4-hydroxycinnamic acid) matrix. Samples were analyzed in positive ion mode on a Voyager-DE STR mass spectrometer (Perseptive Biosystems). Internal sample calibration was achieved with a mixture of insulin, thioredoxin, and myoglobin standards.

Preparation of NMR samples

Uniformly ^{15}N isotopically-labeled DBD (451-694 and 497-694 DBD without a His-tag) was produced by expression in minimal media containing 6.7 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1.5 g/L NaCl, 2 g/L glucose, 10 mL/L Basal Medium Eagle Vitamin solution (GibcoBRL), 162.2 μg/L $FeCl_3$, 2.86 mg/L H_3BO_4, 15 mg/L $CaCl_2 \cdot 2H_2O$, 40 μg/L $CoCl_2 \cdot 6H_2O$, 200 μg/L $CuSO_4 \cdot 5H_2O$, 208 mg/L $MgCl_2 \cdot 6H_2O$, 2 μg/L MoO_3, 208 μg/L $ZnCl_2$, and 1.5 g/L $(^{15}NH_4)_2SO_4$. Growth and purification was as described above except for a 7 hour induction time with IPTG, yielding typically 10-15 mg protein per liter medium. The 451-694 DBD was concentrated to 400 μM in 50 mM potassium phosphate, pH 7.0, 50 mM NaCl, 0.02% NaN_3, 2 mM DTT-d_{10}, and 10% D_2O. Concentrations of this domain above 400 μM proved to be insoluble. The 497-694 DBD was concentrated to 700 μM in 50 mM imidazole-d_4, pH 7.0, 150 mM NaCl, 100 mM Na_2SO_4, 0.02% NaN_3, 2 mM DTT-d_{10}, and 10% D_2O.

NMR methods

NMR data were collected at 20°C or 25°C on a Varian Unity Inova 600 MHz spectrometer. ^{1}H-^{15}N HSQC spectra were obtained with 2048 points and 128 t_1 increments using a gradient sensitivity-enhanced pulse sequence (26). Spectra were processed with the NMRPipe/NMRDraw programs using a cosine apodization function and one round of zero filling (27).

Modified photoactive DNA oligonucleotides

The 5-iodo-2'deoxyuridine-containing single-stranded DNA oligonucleotides (Operon, see table 1) were resuspended in 10 mM triethylammonium acetate buffer, pH 6.0, and purified by acetonitrile gradient on a semipreparative C4 reversed-phase column at 4 mL/min (Vydac). Solutions of the purified oligonucleotides were prepared in deionized water and stored at –20°C. Purity of the DNA was determined to be >99% by MALDI mass spectrometry obtained in negative ion mode using HPA (hydroxypicolinic acid) as the crystallization matrix.

Protein-DNA photocrosslinking

Crosslinking reactions (500 μL of 100 μM protein and DNA) were performed in 50 mM potassium phosphate buffer, pH 7.0, 50 mM NaCl, and 1 mM DTT. The reactions were transferred to a 1 mL polymethylmethacrylate cuvette with a 1 cm path length and irradiated at 325 nm with stirring for 3 hours by an Omnichrome Series 74 He-Cd laser operating at 25-27 mW. Reactions were analyzed by SDS-PAGE.

Equilibrium binding assays by gel shift and filter-binding

The 11mer dGTGTGGGTGTG was 5'-end labeled using T4 DNA kinase according to the Gibco/BRL protocol, with 5 μM DNA and 150 mCi/ml γ^{32}P-ATP. A 25

μL labeling reaction was incubated at 37°C for 30 minutes. Unincorporated ^{32}P was removed using microspin G25 columns (Pharmacia). All assays were conducted in 5 mM HEPES, pH 7.8, 75 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM Na_2EDTA, 1 mM DTT, and 0.1 mg/mL BSA. Equilibrium binding reactions were performed with ^{32}P-11mer at concentrations 10-fold below the dissociation constant and serial dilutions of protein. The reactions were incubated on ice for 30 to 60 minutes to equilibrate. For gel shift assays, 5 μL of each reaction with a small amount of bromophenol blue tracking dye were loaded on a 20 cm x 20 cm x 1.5 mm, 5% acrylamide, nondenaturing gel. Gels were equilibrated at a constant 200 V for 30 to 45 minutes before the samples were loaded. Gels were dried and visualized by Phosphoimager (Molecular Dynamics). For filter binding, 80 μL of each binding reaction was filtered through a 96-well MultiScreen MAHA N4550 filter plate using a MultiScreen Resist Vacuum Manifold (Millipore). The wells were prewashed with 80 μL of binding buffer without BSA, washed 2 x 200 μL after the samples had been filtered, and the filter allowed to dry before being exposed to a Phosphoimager screen. For both assays, spots were quantified (Imagequant) and plots were normalized and fit with a standard two-state binding model: $y=(y_{max})/(1+(K_d/x))$ where x is the concentration of protein and y is the fraction of DNA bound. Equilibrium dissociation constants (K_ds) are reported as an average value plus or minus standard deviation of at least three measurements determined on different days or with different protein preparations.

Off rates measured by native PAGE

Protein (5.55 nM) was incubated for 60 minutes on ice with 555 pM ^{32}P-labeled DNA in 200 μL binding buffer containing bromophenol blue dye. 18 μL aliquots were

removed over a time course; 2 μL of 1 μM unlabeled DNA was added to each aliquot (final concentrations of 5 nM protein, 500 pM ^{32}P-labeled DNA, and 100 nM unlabeled DNA). Time points were analyzed by 5% native acrylamide gel run at 200 V. Plots of protein-bound ^{32}P-labeled DNA versus time were fit to single exponential decay curves: $y = C1(\exp(-k_{off} * x)) + C2$ where x is time in hours, y is the fraction of DNA bound, C1 is the span (Y_{max} - C2), and C2 is the asymptote or plateau of non-specific binding.

Binding titrations to determine the fraction of active protein

^{32}P-labeled DNA (1 nM) and unlabeled DNA (100 nM) were mixed and heated to 90°C for ten minutes and cooled quickly on ice. This mixture was incubated on ice for 1.5 hours with varying amounts of protein (full-length Cdc13, the 451-694 DBD, or the 497-694 DBD) ranging from 0 to 3 molar equivalents. The samples were analyzed by 5% native acrylamide gel.

RESULTS

Limited proteolysis reveals the minimal Cdc13 DNA-binding domain (DBD)

Deletion mapping and limited proteolysis were used previously to map a DNA-binding domain (451-694 DBD) within the 924 amino-acid full-length Cdc13 protein (20,21). This domain has a strong tendency to precipitate at high concentration which is somewhat alleviated at reduced temperatures (15-20°C). The ^{1}H-^{15}N HSQC spectrum of the 451-694 DBD (Figure 1a) reveals that while a folded species is present with well-dispersed chemical shifts in both proton and nitrogen dimensions, the spectrum is dominated by crosspeaks with chemical shifts clustering between 8 and 9 ppm in the proton dimension. The lack of chemical shift dispersion and the presence of sharp linewidths in these peaks indicate the presence of an unfolded species. The features of

this spectrum could be due to an unfolded region of polypeptide in an otherwise folded domain or the presence of both folded and unfolded proteins at equilibrium.

To test the possibility of an unfolded region in an otherwise folded domain, the 451-694 DBD was subjected to limited trypsin digestion at room temperature. Figure 2 illustrates the timecourse of the reaction analyzed by SDS-PAGE. Upon exposure to trypsin, the 28 kD 451-694 DBD almost immediately formed a smaller, 22 kD stable fragment which was remarkably resistant to further cleavage. This reaction pattern did not change in the presence of 1 molar equivalent of the single-stranded DNA ligand dGTGTGGGTGTG, indicating that the unfolded region did not become structured upon binding DNA (data not shown).

MALDI mass spectrometry was used to specifically determine the boundaries of the smaller, stable domain. An identical trypsin digest product was obtained in low salt conditions (10 mM Tris-HCl), which facilitated direct analysis of the reaction. The major product by MALDI (Figure 3) is shown to be a MH(+1) of 21,985 ± 72 Daltons. Examination of the map of predicted trypsin cut sites reveals that this fragment corresponds to one of two predicted fragments: amino acids 504-692 (MH(+1) 21,971) or amino acids 496-685 (MH(+1) 21,979). This result indicates that approximately 50 amino acids at the N-terminus of the 451-694 DBD are particularly susceptible to trypsin cleavage, presumably because this region is unfolded.

The 497-694 DBD forms a stable structural domain

Several shorter candidates of the domain were subcloned for recombinant expression in *E. coli*. A construct of the domain comprising amino acids 497-694 expressed in high yield as a soluble protein. This domain exhibited several favorable

features relative to the 451-694 DBD. It was considerably more soluble than the longer domain and remained soluble at higher temperatures (25-30°C). A comparison of the ^{1}H-^{15}N HSQC spectrum of the parent domain (451-694 DBD) and of the minimal domain (497-694 DBD) is shown in Figure 1. The NMR spectrum was dramatically improved by the removal of the unfolded region. The dispersed resonances of the folded species are almost identical between the two domains. However, the random-coil resonances (Figure 1a) have completely disappeared (Figure 1b), consistent with removal of a large region of unfolded polypeptide that does not affect the structure of the folded domain.

The 497-694 DBD contacts the entire minimal 11-nucleotide DNA

The minimal DNA required for high-affinity binding by full-length Cdc13 is an 11-mer, dGTGTGGGTGTG. This sequence of DNA is complementary to the yeast telomerase RNA template and is representative of yeast telomeric sequence (20,28). Four variants of this minimal DNA were used for photocrosslinking, with the chromophore 5-iodouracil substituted for each of the four thymine bases of the molecule (Table 1). The iodine atom of 5-iodouracil is approximately the same size as the methyl group of thymine; therefore, this substitution is not likely to perturb binding of the protein/DNA complex. Indeed, the K_d of each of the substituted DNAs for binding to the 497-694 DBD was identical to the unsubstituted DNA (data not shown). The long wavelength of 5-iodouracil's absorption and photocrosslinking chemistry (325 nm) disfavors nonspecific excitation of the DNA and protein chromophores (29). As expected, no covalent products were formed and protein degradation was not observed upon irradiation of the 497-694 DBD complexed with unsubstituted DNA (Figure 4a). Study of the timecourse of photocrosslinking revealed that each substituted DNA formed a specific covalent

adduct consistent with 1:1 stoichiometry of the protein and DNA (Figures 4b-4e). The yields of crosslinked species determined by densitometry ranged from 20-40% (Table 1), which are typical for systems using this chromophore (29-31). Preliminary proteolytic digestion of the crosslinked species indicated that the first substituted thymine in the DNA sequence crosslinked to multiple peptides in the protein, whereas the other substituted DNAs crosslinked to a single peptide corresponding to a fragment near the N-terminus of the domain. Identification of the exact site of crosslinking on the protein was precluded by the inability to generate complete proteolytic digests of the protein-DNA adduct. However, we have recently obtained more detailed information on the binding interface of the protein by NMR structural analysis using experiments that measure intermolecular contacts (25). Several aromatic amino acids are located along the interface that would be expected to form crosslinks to 5-iodouracil. These data indicate that the 497-694 DBD contains all the contacts needed for binding the minimal DNA 11-mer.

The 451-694 DBD and the 497-694 DBD bind telomeric DNA more tightly than the full-length protein

To test whether the smaller 497-694 DBD was both necessary and sufficient for function, gel shift and filter binding assays were used to determine the equilibrium binding dissociation constant (K_d) of this domain with the single-stranded telomeric DNA substrate dGTGTGGGTGTG. Figure 5 compares the binding curves determined by filter-binding assays of full-length Cdc13, the 451-694 DBD, and the 497-694 DBD. Dissociation constants (K_ds) measured by gel-shift assay yielded the same results (data not shown). The measured K_d of 310 ± 50 pM for full-length Cdc13 confirms the result

of previous studies (14,20). However, we observed that the K_d for 451-694 DBD was 5 ± 1 pM, which is substantially lower than previously reported - 370 pM (20) and ranging from 72-240 nM for similar substrates (32). The K_d for the 497-694 DBD (both with and without the His-tag) was 3 ± 1 pM, or approximately 100-fold tighter than for full-length Cdc13 and similar to that of the 451-694 DBD. Binding titrations well above the K_d indicated that protein preparations were over 75% active under these conditions with a binding stoichiometry of 1:1 (data not shown). Therefore the differences that we observed in binding were not due to a difference in the fraction of active protein.

The dissociation rates of the three proteins are similar

Dissociation rates were measured by competition assay for the three proteins and are presented in Figure 6. These dissociation rates are quite similar, with rate constants of 2.8×10^{-4} min^{-1} for full-length Cdc13, 3.3×10^{-4} min^{-1} for 451-694 DBD, and 4.3×10^{-4} min^{-1} for 497-694 DBD. This gives a half-life for the complex of approximately 41 hours for full-length Cdc13, 27 hours for the 497-694 DBD, and 35 hours for the 451-694 DBD. Based on the similarity of the dissociation rates, the large difference in K_d between full-length protein and the two domains must be due to an association rate effect. Association rates were too fast to measure accurately by gel shift assay directly. From the measured K_ds and dissociation rates, the calculated association rates are as follows: 9.0×10^{5} $M^{-1}min^{-1}$ for full-length Cdc13, 6.6×10^{7} $M^{-1}min^{-1}$ for the 451-694 DBD, and 1.4×10^{8} $M^{-1}min^{-1}$ for the 497-694 DBD.

DISCUSSION

Single-stranded nucleic acids are involved in a wide array of fundamental biological functions, including telomere regulation, DNA replication, repair and

recombination, transcriptional mechanisms, translation, and RNA splicing (33-35). Recognition of single-stranded nucleic acids has been implicated in many pathological processes in humans, ranging from cancer and aging to various infectious diseases. Therefore the sequence-specific and non-specific recognition of single-stranded nucleic acids by proteins is crucial for maintaining, manipulating, and utilizing the genetic material contained in cells. Relatively little is known concerning the requirements, either structural or functional, for sequence recognition in a single-stranded context. Cdc13 is an essential protein in *S. cerevisiae* that regulates telomere capping and telomeric replication (36). The single-stranded telomeric DNA-binding domain is central to these functions and can substitute for full-length protein when fused to appropriate binding partners (16,18). The high affinity and sequence specificity of single-stranded DNA recognition exhibited by Cdc13 (14,20) make this an ideal system to further our understanding of sequence-specific single-stranded DNA binding.

Previous identification of an independent DNA-binding domain of Cdc13 (451-694 DBD) has facilitated biochemical studies of its single-stranded DNA-binding activity (20,32). Although the DBD defined by amino acids 451-694 is competent for single-stranded telomeric DNA binding *in vitro* and can function *in vivo*, the NMR data presented here clearly show that it is not a completely folded domain. The ^{1}H-^{15}N HSQC fingerprint spectrum of this domain (Figure 1a) contains several intense, poorly dispersed resonances indicative of an unstructured state superpositioned on the well-dispersed resonances indicative of a folded protein. Protein domains involved in a diverse array of cellular processes have been found to be intrinsically unstructured and only fold upon binding to their protein or nucleic acid partners (37). Thus, we considered the possibility

that the unfolded region is directly involved in contacting DNA. However, in this case the unfolded region does not fold upon binding and does not affect single-stranded telomeric DNA binding by Cdc13 directly. The 451-694 domain also exhibits poor solubility and is therefore unsuitable for *in vitro* characterization. Presumably, a construct of this domain expressed in yeast may have unexpected, variable effects as well.

In this work, the Cdc13 DBD was refined by limited trypsin digestion and MALDI mass spectrometry (Figures 2 and 3), producing a domain (497-694 DBD) that is both structurally and functionally independent. This represents the true, minimal DNA-binding domain. In contrast to the 451-694 DBD, the ^{1}H-^{15}N HSQC spectrum of the 497-694 DBD reveals the presence of a completely folded species (Figure 1b). Careful examination of the spectrum obtained on the 451-694 DBD reveals that the well-dispersed resonances of the 497-694 DBD are present within the spectrum of the 451-694 DBD, clearly demonstrating that the folded region is also present in the longer domain. Further biochemical characterization using affinity photocrosslinking revealed that the smaller, folded domain contacts the entire 11-mer of single-stranded DNA required for high affinity and specificity binding (20). Affinity photocrosslinking has previously been used successfully as a probe of contacts between protein and single-stranded DNA (38). In particular, crosslinks found between chromophore-containing DNA and the *Oxytricha nova* telomere end binding protein corresponded to sites of protein/DNA contacts in the high resolution structure (22,29,39). Similarly high-yielding protein-DNA crosslinks were obtained in this system with use of the 5-iodouracil chromophore. We have shown that sites throughout the DNA 11-mer crosslink and therefore interact with the 198 amino

acid minimal binding domain defined here. This result has been subsequently confirmed by mapping the protein/DNA interface using NMR spectroscopy, which revealed that several amino acids capable of crosslinking do contact DNA (25).

Interestingly, both DBDs bind DNA with tighter affinity ($K_d \approx$ 3-5 pM) than the full-length protein (K_d = 310 pM) (Figure 5). Thus the smaller, 497-694 DBD contains the essential region for contacting single-stranded telomeric DNA. The difference in affinity between the isolated domains and full-length protein could simply be an artifact of extracting the domain from the full-length protein. However, enhanced binding of functional domains relative to full-length protein has been observed in the other telomere end-binding proteins *O. nova* TEBP α subunit (40) and *S. pombe* Pot1p (6), perhaps indicating that this is a general feature of this family of telomere end-binding proteins. The full-length protein may contain a region inhibitory to binding which could play a regulatory role *in vivo* by attenuating the extremely tight binding of the DBD. In the case of Pot1 protein, deletion of a large COOH-terminal segment increased binding affinity by approximately 10-fold, while in *O. nova* TEBP α subunit several different truncations at the COOH-terminus increased binding. Additionally, several splice variants of human Pot1 protein (hPot1) have been identified and appear to interact differentially with single-stranded human telomeric DNA (P. Baumann, E. Podell, and T. R. Cech, personal communication).

Our characterization of the 451-694 DBD has generated results that are in contrast to previous studies of the same domain (20,32). In our study we have used highly purified, soluble preparations of the DBDs, taking care to ensure that over 75% of the protein was active. We have observed that it is critical to include a small amount of BSA

(or Nonidet P-40 detergent) in the binding reactions to prevent loss of the protein to surfaces and aggregation or precipitation of the protein. When this protocol was not followed we obtained low and irreproducible binding, presumably due to non-specific loss of protein and/or protein inactivation (data not shown). Previous studies intended to determine the binding affinity of the 451-694 DBD were not performed in the presence of BSA or detergent (20; 32; V. Lundblad, personal communication). This may explain the discrepancy between previous results and the binding affinities reported here.

The rates of dissociation of the DNA ligand for full-length Cdc13, 451-694 DBD, and 497-694 DBD are uniformly slow, yet similar. The calculated rate of association of the 451-694 DBD and the 497-694 DBD with DNA is 2-3 orders of magnitude faster than for full-length protein. Thus, the differences in binding affinity are primarily due to differences in the rates of association. We have not determined *in vitro* if the attenuation of binding in the full-length protein is due to regions NH_2- or COOH-terminal to the DBD. Further study of this phenomenon is underway to determine what effect this attenuation has on yeast telomeres *in vivo*.

Cdc13 performs critical functions with partner protein complexes both in capping the telomere, protecting it from degradation and fusion, and in regulating telomeric replication and length. The single-stranded telomeric DNA-binding domain of Cdc13 is central to its function and exhibits unusually high affinity and specificity for its DNA target. We have delineated the true minimal folded domain that is both necessary and sufficient for high-affinity binding to single-stranded telomeric DNA. The present work provides insights into this important mode of DNA recognition.

ACKNOWLEDGMENTS

We would like to gratefully acknowledge and thank members of the Wuttke lab, especially Douglas Theobald for performing titration binding experiments, Thorsten Schäfer for contributions to the photocrosslinking studies, and Corey Mandel for subcloning the 497-694 DBD his-tagged construct. We thank Rachel Mitton-Fry, Leslie Glustrom, Art Pardi, and Fiona Jucker for careful reading of the manuscript. We thank Vicki Lundblad and Timothy Hughes for access to data before publication, for plasmids encoding both full-length Cdc13 and the 451-694 DBD, and for SF9 cell pellets containing full-length Cdc13. We are thankful for all sources of funding for this research, which has been provided by the NIH (GM59414), the American Cancer Society, a University of Colorado Junior Faculty Development Award, and a predoctoral fellowship from the U.S. Army Breast Cancer Research Program (EMA). The NMR instrumentation was purchased with partial support from NIH RR11969 and NSF 9602941. We also thank the W. M. Keck Foundation for support of the Molecular Structure Program on the Boulder campus.

REFERENCES

1. Shore, D. (2001) Telomeric chromatin: replicating and wrapping up chromosome ends. *Curr. Opin. Genet. Dev.*, **11**, 189-198.

2. Blackburn, E. H. (2001) Switching and signaling at the telomere. *Cell*, **106**, 661-673.

3. de Lange, T. (2001) Telomere capping - one strand fits all. *Science*, **292**, 1075-1076.

4. Gottschling, D. E. and Zakian, V. A. (1986) Telomere proteins: specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell*, **47**, 195-205.

5. Price, C. M. and Cech, T. R. (1987) Telomeric DNA-protein interactions of *Oxytricha* macronuclear DNA. *Genes Dev.*, **1**, 783-793.

6. Baumann, P. and Cech, T. R. (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science*, **292**, 1171-1175.

7. Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H. and de Lange, T. (1999) Mammalian telomeres end in a large duplex loop. *Cell*, **97**, 503-514.

8. Weinert, T. A. and Hartwell, L. H. (1993) Cell cycle arrest of *cdc* mutants and specificity of the RAD9 checkpoint. *Genetics*, **134**, 63-80.

9. Garvik, B., Carson, M. and Hartwell, L. (1995) Single-stranded DNA arising at telomeres in *cdc13* mutants may constitute a specific signal for the *RAD9* checkpoint. *Mol. Cell. Biol.*, **15**, 6128-6138.

10. Diede, S. J. and Gottschling, D. E. (1999) Telomerase-mediated telomere addition *in vivo* requires DNA primase and DNA polymerases alpha and delta. *Cell*, **99**, 723-733.

11. Diede, S. J. and Gottschling, D. E. (2001) Exonuclease activity is required for sequence addition and Cdc13p loading at a *de novo* telomere. *Curr Biol.*, **11**, 1336-1340.

12. Booth, C., Griffith, E., Brady, G. and Lydall, D. (2001) Quantitative amplification of single-stranded DNA (QAOS) demonstrates that *cdc13-1* mutants generate ssDNA in a telomere to centromere direction. *Nucleic Acids Res.*, **29**, 4414-22.

13. Chandra, A., Hughes, T. R., Nugent, C. I. and Lundblad, V. (2001) Cdc13 both positively and negatively regulates telomere replication. *Genes Dev.*, **15**, 404-414.

14. Nugent, C. I., Hughes, T. R., Lue, N. F. and Lundblad, V. (1996) Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science*, **274**, 249-252.

15. Lingner, J., Cech, T. R., Hughes, T. R. and Lundblad, V. (1997) Three Ever Shorter Telomere (*EST*) genes are dispensible for *in vitro* yeast telomerase activity. *Proc. Natl. Acad. Sci. USA*, **94**, 11190-11195.

16. Pennock, E., Buckley, K. and Lundblad, V. (2001) Cdc13 delivers separate complexes to the telomere for end protection and replication. *Cell*, **104**, 387-396.

17. Lin, J.-J. and Zakian, V. A. (1996) The *Saccharomyces CDC13* protein is a single-strand $TG_{1\text{-}3}$ telomeric DNA-binding protein *in vitro* that affects telomere behavior *in vivo*. *Proc. Natl. Acad. Sci. USA*, **93**, 13760-13765.

18. Evans, S. K. and Lundblad, V. (1999) Est1 and Cdc13 as comediators of telomerase access. *Science*, **286**, 117-120.

19. Hughes, T. R., Evans, S. K., Weilbaecher, R. G. and Lundblad, V. (2000) The Est3 protein is a subunit of yeast telomerase. *Curr Biol.*, **10**, 809-812.

20. Hughes, T. R., Weilbaecher, R. G., Walterscheid, M. and Lundblad, V. (2000) Identification of the single-strand telomeric DNA binding domain of the *Saccharomyces cerevisiae* Cdc13 protein. *Proc. Natl. Acad. Sci. USA*, **97**, 6457-6462.

21. Wang, M.-J., Lin, Y.-C., Pang, T.-L., Lee, J.-M., Chou, C.-C. and Lin, J.-J. (2000) Telomere-binding and Stn1p-interacting activities are required for the essential function of *Saccharomyces cerevisiae* Cdc13p. *Nucleic Acids Res.*, **28**, 4733-4741.

22. Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A. and Schultz, S. C. (1998) Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell*, **95**, 963-974.

23. Horvath, M. P. and Schultz, S. C. (2001) DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex. *J. Mol. Biol.*, **310**, 367-377.

24. Classen, S., Ruggles, J. A. and Schultz, S. C. (2001) Crystal structure of the N-terminal domain of *Oxytricha nova* telomere end-binding protein alpha subunit both uncomplexed and complexed with telomeric ssDNA. *J. Mol. Biol.*, **314**, 1113-1125.

25. Mitton-Fry, R. M., Anderson, E. M., Hughes, T. R., Lundblad, V. and Wuttke, D. S. (2002) Conserved structure for single-stranded telomeric DNA recognition. *Science*, **296**, 145-147.

26. Kay, L. E., Keifer, P. and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.*, **114**, 10663-10665.

27. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277-293.

28. Singer, M. S. and Gottschling, D. E. (1994) TLC1: template RNA component of *Saccharomyces cerevisiae* telomerase. *Science*, **266**, 404-409.

29. Willis, M. C., Hicke, B. J., Uhlenbeck, O. C., Cech, T. R. and Koch, T. H. (1993) Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science*, **262**, 1255-1257.

30. Steen, H., Petersen, J., Mann, M. and Jensen, O. N. (2001) Mass spectrometric analysis of a UV-cross-linked protein-DNA complex: tryptophans 54 and 88 of *E. coli* SSB cross-link to DNA. *Protein Sci.*, **10**, 1989-2001.

31. Wong, D. L. and Reich, N. O. (2000) Identification of tyrosine 204 as the photo-cross-linking site in the DNA-*Eco*RI DNA methyltransferase complex by electrospray ionization mass spectrometry. *Biochemistry*, **39**, 15410-15417.

32. Lin, Y.-C., Hsu, C.-L., Shih, J.-W. and Lin, J.-J. (2001) Specific binding of single-stranded telomeric DNA by Cdc13p of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **276**, 24588-24593.

33. Swamynathan, S. K., Nambiar, A. and Guntaka, R. V. (1998) Role of single-stranded DNA regions and Y-box proteins in transcriptional regulation of viral and cellular genes. *Faseb J.*, **12**, 515-522.

34. Broomfield, S., Hryciw, T. and Xiao, W. (2001) DNA postreplication repair and mutagenesis in *Saccharomyces cerevisiae*. *Mutat Res.*, **486**, 167-184.

35. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. and Yokoyama, S. (1999) Structural basis for the recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature*, **398**, 579-585.

36. Lustig, A. J. (2001) Cdc13 subcomplexes regulate multiple telomere functions. *Nature Struct. Biol.*, **8**, 297-299.

37. Wright, P. E. and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, **293**, 321-331.

38. Meisenheimer, K. M. and Koch, T. H. (1997) Photocross-Linking of Nucleic Acids to Associated Proteins. *Critical Reviews in Biochemistry and Molecular Biology*, **32**, 101-140.

39. Hicke, B. J., Willis, M. C., Koch, T. H. and Cech, T. R. (1994) Telomeric Protein-DNA Contacts Identified by Photo-Cross-Linking Using 5-Bromodeoxyuridine. *Biochemistry*, **33**, 3364-3373.

40. Fang, G., Gray, J. T. and Cech, T. R. (1993) *Oxytricha* telomere-binding protein: separable DNA-binding and dimerization domains of the α-subunit. *Genes Dev.*, **7**, 870-882.

FIGURE LEGENDS

Figure 1. Comparison of Cdc13 DBD ^{1}H-^{15}N HSQC spectra. a. ^{1}H-^{15}N HSQC spectrum of 400 μM 451-694 DBD at 600 MHz, 20°C. Sample was prepared in 50 mM potassium phosphate, pH 7.0, 50 mM NaCl, 0.02% NaN_3, 2 mM DTT-d_{10}, and 10% D_2O. b. ^{1}H-^{15}N HSQC spectrum of 700 μM 497-694 DBD at 600 MHz, 25°C. Sample was prepared in 50 mM imidazole-d_4, pH 7.0, 150 mM NaCl, 100 mM Na_2SO_4, 0.02% NaN_3, 2 mM DTT-d_{10}, and 10% D_2O.

Figure 2. Timecourse of limited trypsin digestion of the 451-694 DBD. Reactions were performed as described in Materials and Methods. Lane m, protein markers; other lanes in minutes of time of the reaction.

Figure 3. MALDI-TOF mass spectrum of trypsin digest products. a, minor digestion product; b-e, protein internal calibration standards as follows: b, insulin +1; c, myoglobin +2; d, thioredoxin +1; e, myoglobin +1; f, major digestion product; g, trypsin.

Figure 4. 15% SDS gel of photocrosslinked 497-694 DBD/DNA at 100 μM each. Shown are time courses for irradiation at 325 nm, with lanes labeled in minutes. Conditions are stated in Materials and Methods. a. WT DNA of sequence dGTGTGGGTGTG, b. DNA1, c. DNA2, d. DNA3, e. DNA4.

Figure 5. Fraction of DNA bound as a function of protein concentration used to determine the equilibrium dissociation binding constant. ◆, full-length Cdc13, ■, 451-

694 DBD, and ●, 497-694 DBD bound to dGTGTGGGTGTG. Each curve is an average of at least three separate experiments conducted by filter binding. Plots were fit with a standard two-state binding model. Equilibrium dissociation constants (K_ds) are 310 ± 50 pM, 5 ± 1 pM, and 3 ± 1 pM, respectively. Full-length protein is His-tagged, 451-694 DBD is not His-tagged, and 497-694 DBD was measured with and without a His-tag (no difference in binding).

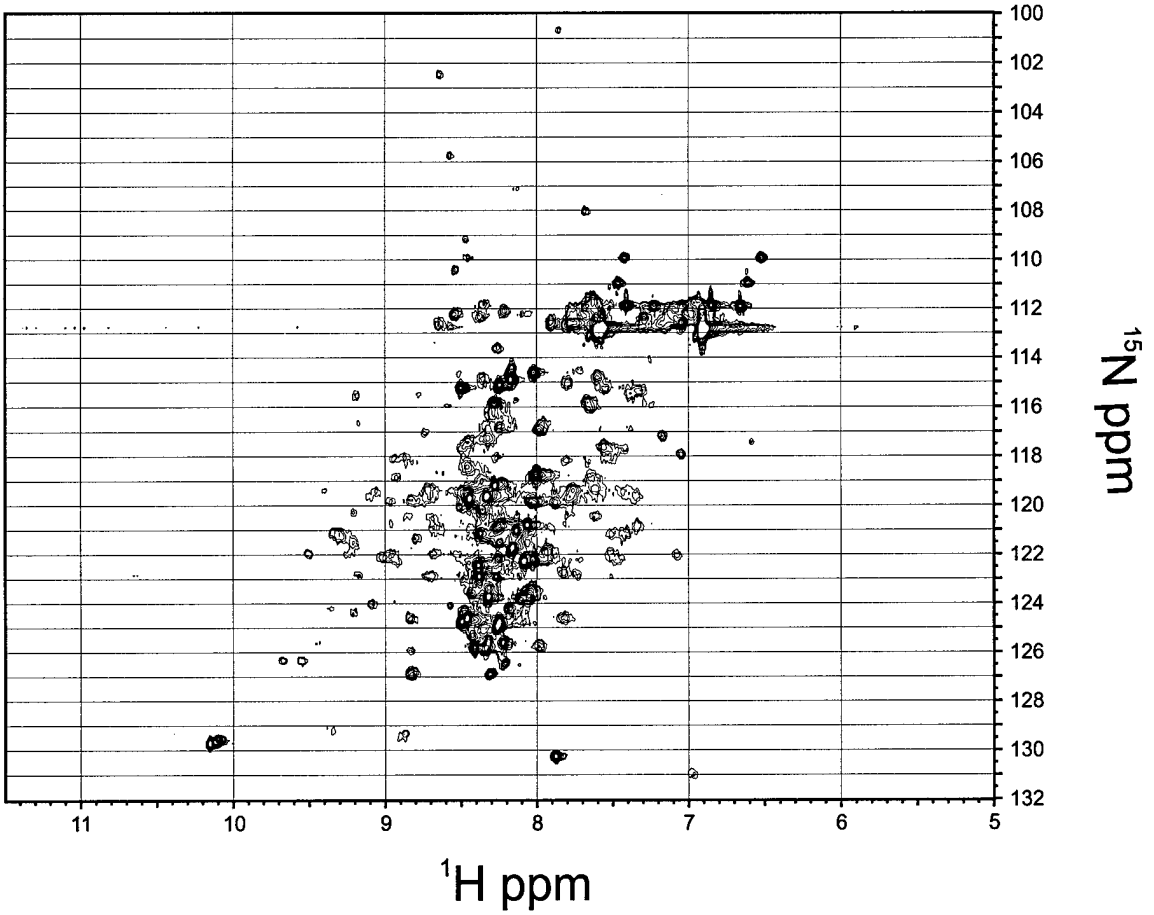
Figure 6. Fraction of labeled DNA bound as a function of time as measured by native PAGE, used to determine the dissociation rate constant. ◆, full-length Cdc13, ■, 451-694 DBD, and ●, 497-694 DBD. Experiments were conducted as stated in Materials and Methods. Full-length protein is COOH-terminally His-tagged, while both domains are not His-tagged.

Table 1. Yields of crosslinked photoactive DNAs

Name	Sequence	% yield
WT	dGTGTGGGTGTG	0
DNA1	dG^{I}UGTGGGTGTG	19
DNA2	dGTG^{I}UGGGTGTG	22
DNA3	dGTGTGGG^{I}UGTG	40
DNA4	dGTGTGGGTG^{I}UG	19

^{I}U represents 5-iodouracil substituted for the base thymine

a


100
102
104
106
108
110
112
114
116
118
120
122
124
126
128
130
132
11
10
9
8
7
6
5
^{15}N ppm
^{1}H ppm

b

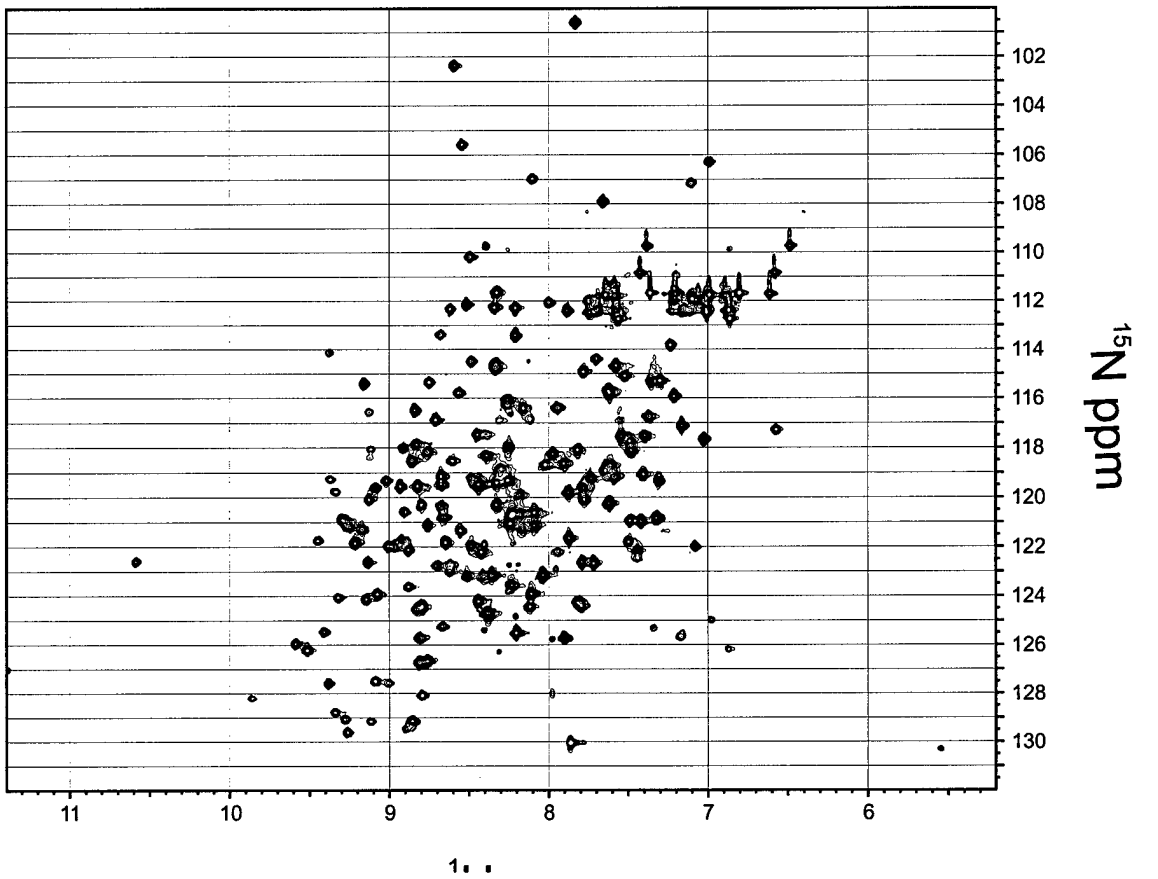

102
104
106
108
110
112
114
116
118
120
122
124
126
128
130
11
10
9
8
7
6
^{15}N ppm

Figure 2

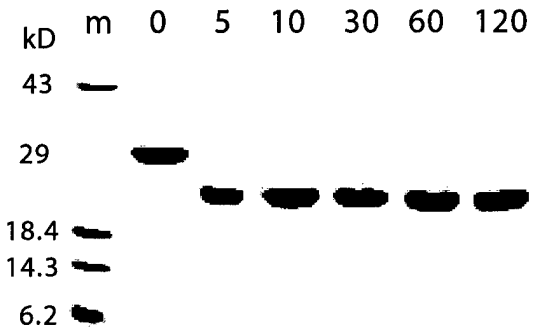


Figure 3

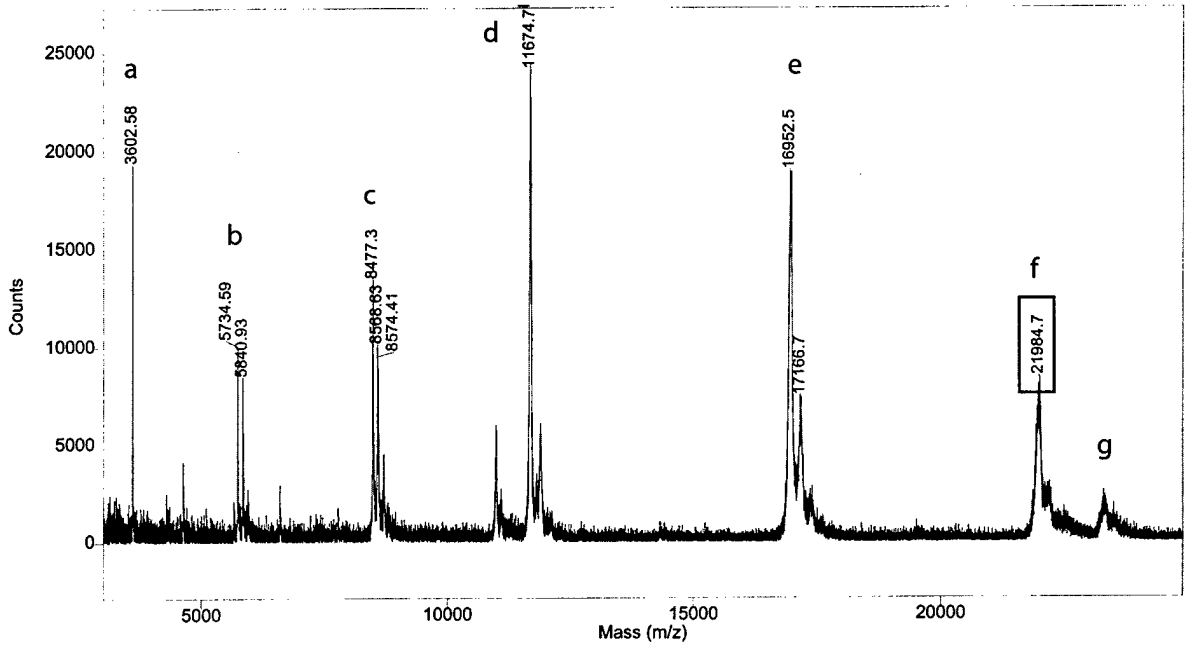

a
b
c
d
e
f
g
3602.58
5734.59
5840.93
8477.3
8566.63
8574.41
11674.7
16952.5
17166.7
21984.7
25000
20000
15000
10000
5000
0
Counts
5000
10000
15000
20000
Mass (m/z)

Figure 4

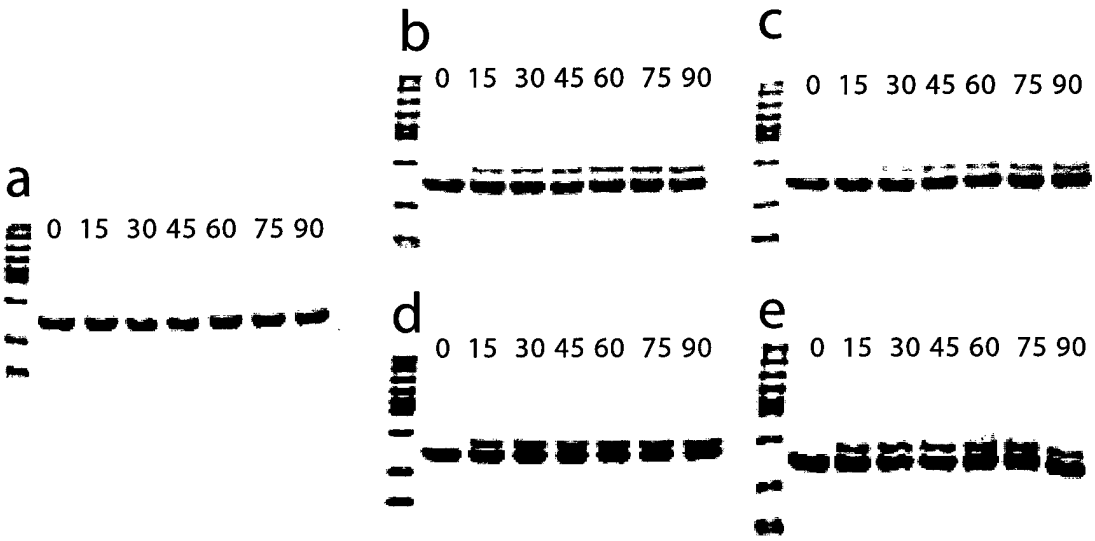

a
0 15 30 45 60 75 90
b
0 15 30 45 60 75 90
c
0 15 30 45 60 75 90
d
0 15 30 45 60 75 90
e
0 15 30 45 60 75 90

Figure 5

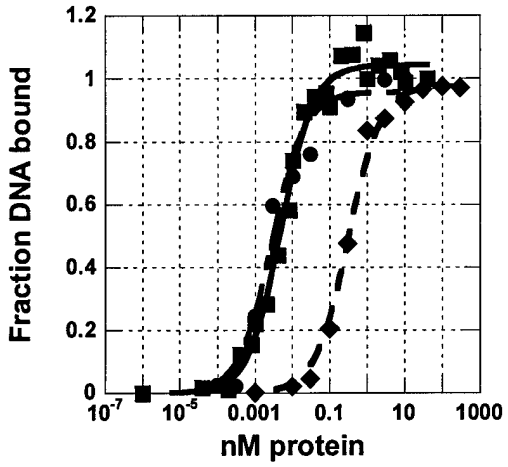
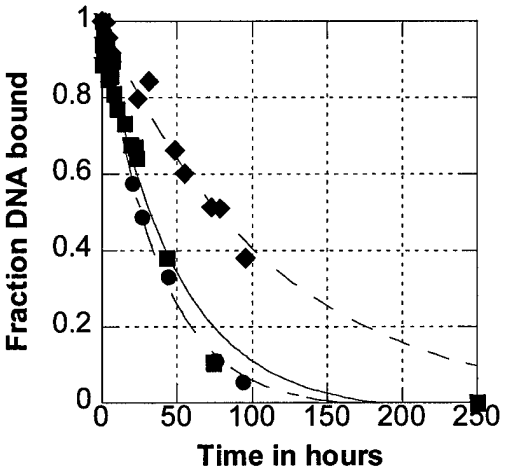

Fraction DNA bound
1.2
1
0.8
0.6
0.4
0.2
0
10^{-7}
10^{-5}
0.001
0.1
10
1000
nM protein

Figure 6


Fraction DNA bound
1
0.8
0.6
0.4
0.2
0
0
50
100
150
200
250
Time in hours

Tagawa, T. Tomoya, K. Nomura, T. Hanagaki, and T. Matsuyama. We thank T. Takase and M. Matsui for sampling of the dark-grown plants. Supported by a grant for Genome Research from RIKEN; the Program for Promotion of Basic Research Activities for Innovative Biosciences; the Special Coordination Fund of the Science and Technology Agency; and a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan to K.S. Also supported in part by a Grant-in-Aid for Scientific Research on Priority Areas (C) "Genome Science" from the Ministry of Education, Science, Sports and Culture of Japan to M.S. and by a Research Grant for the RIKEN Genome Exploration Research Project from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government to Y.H.

19 February 2002; accepted 12 March 2002
Published online 21 March 2002;
10.1126/science.1071006
Include this information when citing this paper.

Conserved Structure for Single-Stranded Telomeric DNA Recognition

Rachel M. Mitton-Fry,[1] Emily M. Anderson,[1] Timothy R. Hughes,[2]* Victoria Lundblad,[2,3] Deborah S. Wuttke[1]†

The essential Cdc13 protein in the yeast *Saccharomyces cerevisiae* is a single-stranded telomeric DNA binding protein required for chromosome end protection and telomere replication. Here we report the solution structure of the Cdc13 DNA binding domain in complex with telomeric DNA. The structure reveals the use of a single OB (oligonucleotide/oligosaccharide binding) fold augmented by an unusually large loop for DNA recognition. This OB fold is structurally similar to OB folds found in the ciliated protozoan telomere end-binding protein, although no sequence similarity is apparent between them. The common usage of an OB fold for telomeric DNA interaction demonstrates conservation of end-protection mechanisms among eukaryotes.

Telomeres are the specialized nucleoprotein complexes that cap eukaryotic chromosomes, protecting chromosome ends from unregulated degradation and end-to-end fusion. Telomeric DNA is typically composed of repetitive, noncoding sequence terminating in a single-stranded TG-rich overhang. Several mechanisms have been identified for capping this overhang, ranging from sequestration through protein binding in ciliates and yeasts to t-loop formation in mammals (*1–3*). Proteins that specifically bind to this single-stranded overhang, such as the *Oxytricha nova* telomere end-binding protein (TEBP) (*4*, *5*), the *Schizosaccharomyces pombe* protection of telomeres 1 (Pot1) and human Pot1 (*6*), and the *Saccharomyces cerevisiae* Cdc13 (*7*, *8*), are involved in telomeric end protection. For example, depletion of Cdc13 activity causes extensive resection of the 5′ strand of the yeast telomere and DNA damage–dependent cell cycle arrest (*9–12*), whereas deletion of the *pot1* gene leads to complete telomere loss and cell death (*6*). Cdc13 is also required for telomere elongation as a positive regulator of telomerase (*7*, *13*). Cdc13 is believed to fulfill both of these important, yet disparate, roles through localization to the 3′ single-stranded telomeric end, followed by recruitment of relevant complexes to the telomere through protein-protein interactions (*14–16*).

Evidence for conservation of telomeric end-protection proteins among distantly related eukaryotes has been elusive. Although the Pot proteins were originally identified on the basis of weak sequence similarity to the NH_2-terminal portion of the α subunit of the heterodimeric *O. nova* TEBP (*6*), no similarity was apparent between any of these proteins and Cdc13. To investigate the requirements for telomeric end protection and sequence-specific interaction with single-stranded DNA (ssDNA), we determined the solution structure of the Cdc13 DNA binding domain (DBD) in complex with telomeric ssDNA. This 23.5-kD domain retains DNA binding activity and specificity (*17–19*), and fusions of the DBD with other components of the end-protection or telomerase machinery eliminate the need for full-length protein in vivo (*14*, *15*). The ssDNA 11-nucleotide (nt) oligomer dGTGTGGGTGTG in the complex is the minimal Cdc13 binding site (*17*) and the complement to the center of the coding region of the telomerase RNA template (*20*).

The high-resolution Cdc13 DBD structure in complex with ssDNA (Fig. 1) was calculated from a total of 2865 nuclear magnetic

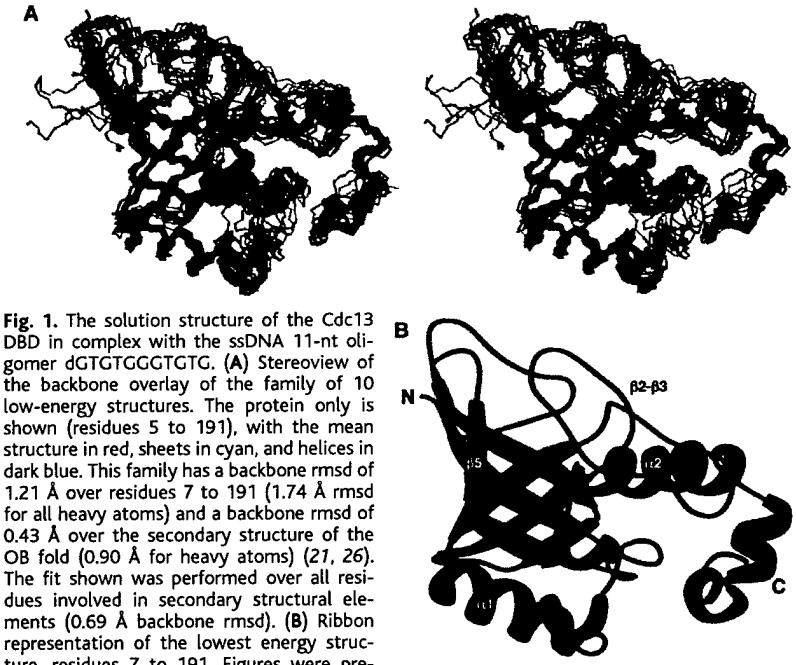


Fig. 1. The solution structure of the Cdc13 DBD in complex with the ssDNA 11-nt oligomer dGTGTGGGTGTG. **(A)** Stereoview of the backbone overlay of the family of 10 low-energy structures. The protein only is shown (residues 5 to 191), with the mean structure in red, sheets in cyan, and helices in dark blue. This family has a backbone rmsd of 1.21 Å over residues 7 to 191 (1.74 Å rmsd for all heavy atoms) and a backbone rmsd of 0.43 Å over the secondary structure of the OB fold (0.90 Å for heavy atoms) (*21*, *26*). The fit shown was performed over all residues involved in secondary structural elements (0.69 Å backbone rmsd). **(B)** Ribbon representation of the lowest energy structure, residues 7 to 191. Figures were prepared with MOLMOL (*33*) and RIBBONS (*34*).

[1]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA. [2]Interdepartmental Program in Cell and Molecular Biology, [3]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA.

*Present address: University of Toronto, Banting and Best Department of Medical Research, Toronto, Ontario M5G 1L6, Canada.
†To whom correspondence should be addressed. E-mail: deborah.wuttke@colorado.edu

resonance (NMR) restraints (*21*). Comparison of the Cdc13 DBD to the structural database unequivocally places Cdc13 in the oligonucleotide-binding superfamily of OB fold proteins (*22*, *23*). The OB fold is a small structural motif used for oligonucleotide, oligosaccharide, and oligopeptide binding (*24*). This fold, exemplified by verotoxin-1, staphylococcal nuclease, and the anticodon-binding domain of asp-tRNA synthetase, cannot yet be predicted on the basis of sequence comparisons. The canonical OB fold, also seen in the Cdc13 DBD, consists of a β barrel formed by two orthogonally packed three-stranded antiparallel β sheets. Sheet 1 is composed of β1, β2, and β3, and sheet 2 comprises β5, β4, and β1. In the Cdc13 DBD, numerous NOEs (nuclear Overhauser effects) between β3 and β5 close the barrel, and an α helix between β3 and β4 caps the bottom of the barrel. The Cdc13 DBD has an unusually long, 30-residue loop between β2 and β3 that packs tightly over the β2 and β3 strands. This loop is structurally well-defined, as indicated by heteronuclear NOE measurements, although it has no regular secondary structure. An α-helical region extends the domain COOH-terminally beyond the OB fold.

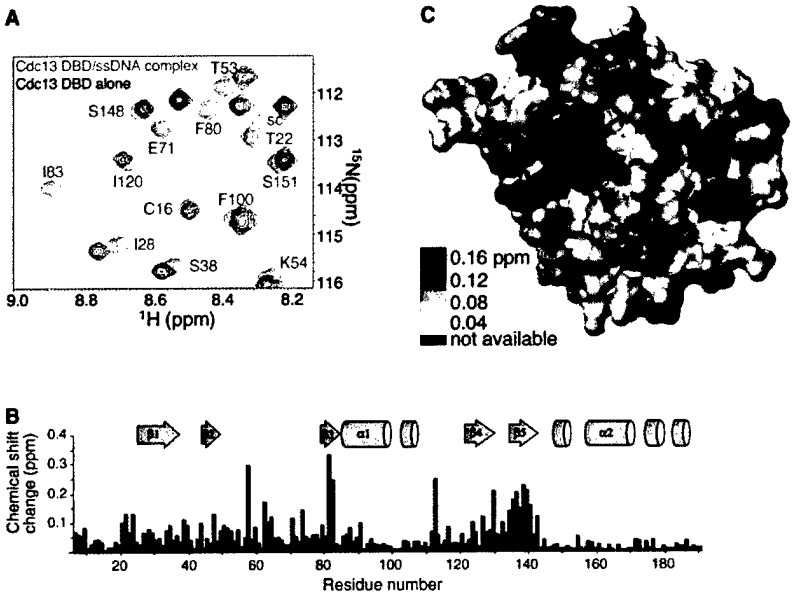


Fig. 2. Interaction of the Cdc13 DBD with single-stranded telomeric DNA determined by chemical shift perturbation. **(A)** Comparison of DBD chemical shifts in the presence and absence of DNA. Overlay of a region of the ^{15}N-^{1}H HSQC (heteronuclear single-quantum coherence) spectra of protein-ssDNA complex (red) and of protein alone (black). Crosspeaks from the protein/DNA complex have been labeled on the spectrum (sc, side chain) (*26*, *35*). Because the complex binds in the slow-exchange time regime, assignments for protein alone cannot be determined by titration. **(B)** Minimal chemical shift perturbation upon DNA binding and secondary structural elements mapped on the protein sequence. Substantial chemical shift changes occur throughout the OB fold portion of the DBD, concentrating in the β-barrel and loop regions. Perturbation values have been calculated according to the equation: perturbation = $\sqrt{(\Delta ppm_{H,min})^2+(0.17\times\Delta ppm_{N,min})^2}$, where $\Delta ppm_{H,min}$ and $\Delta ppm_{N,min}$ are the minimal chemical shift differences (in parts per million) for proton and nitrogen, respectively (*36*). Gray shading indicates residues for which no backbone information is available (predominantly prolines). This method assumes that the crosspeak in the spectrum of protein alone with the least chemical shift change from any given peak in the complex spectrum corresponds to the same residue. Thus, this analysis is an underestimation of the true perturbation upon DNA binding. **(C)** Minimal chemical shift perturbation upon DNA binding mapped on the DBD structure, residues 7 to 191. Yellow to red shading indicates residues with increasing chemical shift perturbation upon DNA binding. Gray shading indicates residues for which no backbone information is available (predominantly prolines).

The DNA binding site on the protein surface was identified by comparison of NMR chemical shifts in the presence and absence of DNA and analysis of intermolecular NOEs between the protein and DNA. DNA binding induced extensive chemical shift changes throughout the OB fold region of the domain and were most pronounced in β3, β4, and β5, and across the loop between β2-β3 (Fig. 2). The COOH-terminal helical region showed no evidence of involvement in DNA binding. We next directly localized the protein-DNA interface on the basis of more than 50 intermolecular NOEs observed between the protein and DNA (Fig. 3) (*25*). These protein-DNA contacts unambiguously define an extensive intermolecular interface that coincides with the binding surface indicated by the chemical shift changes. This interaction surface extends across sheet 2, through the cleft defined by the loops between β3-α1 and β4-β5, over strand β3, and across the length of the long loop between β2-β3. Interestingly, 9 of the 11 nucleotides in the DNA molecule showed NOE contacts to at least one amino acid in the protein. Thirteen residues

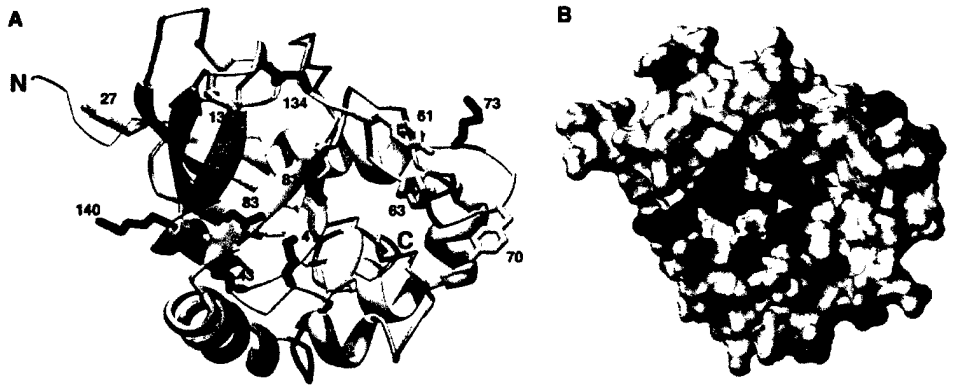


Fig. 3. Interaction of the Cdc13 DBD with single-stranded telomeric DNA as seen by direct NOE contacts. **(A)** Amino acids residues involved in intermolecular NOE contacts with the DNA are mapped on the protein structure (*26*). Tyrosine residues are highlighted in yellow, basic residues in blue, and hydrophobic residues in magenta. **(B)** The above contact residues mapped on the protein surface, shaded as in (A).

Fig. 4. Structural comparison of the Cdc13 DBD OB fold with the NH_2-terminal *O. nova* α OB fold. (**A**) Overlay of the two OB folds. The Cdc13 DBD (residues 10 to 148) is shown in cyan (*26*), and the *O. nova* α OB fold (residues 37 to 150) is shown in gold. Fits were performed with LSQMAN (*37*). (**B**) Comparison of the DNA binding interfaces of the two OB folds. Colors are as described in (A), with contact residues of the Cdc13 DBD OB fold (left) highlighted in red and those of the NH_2-terminal *O. nova* α OB fold (right) in green. This figure was rotated 40° from (A) (the same orientation as seen in Fig. 3) to illustrate the size difference between the two interfaces. *O. nova* α contact residues are taken from (*28*).

were unambiguously identified at the protein-DNA interface, including five aromatic (Y27, Y61, Y63, Y70, and Y131), three hydrophobic (A43, I83, and I138), and five basic amino acids (K41, K73, K81, K134, and R140) (*26*). The predominance of aromatic, hydrophobic, and basic residues in the Cdc13 DBD interface suggests that the aromatic stacking, hydrophobic interactions, and phosphate contacts typically seen in OB fold–ligand structures are also critical for ssDNA binding in the Cdc13 DBD.

OB folds typically interact with a small ligand (e.g., 2 to 5 nucleotides for OB folds that recognize nucleic acids) through interactions with the loops between β1-β2, β3-α, and β4-β5 (*24*). This mode of recognition is also observed in the Cdc13 DBD-ssDNA interaction. However, the Cdc13 DBD markedly expands its interaction surface by using a large loop between β2-β3 for DNA binding (*27*). The extended interface can accommodate the entire DNA molecule, explaining the requirement for at least an 11-nt oligomer of cognate ssDNA for full binding affinity (*17*). This exploitation of the Cdc13 DBD loop for ligand recognition illustrates the substantial malleability and adaptability of the OB fold.

Although the Pot proteins have not yet been structurally characterized, the structure of the ternary complex of the related *O. nova* TEBP bound to a 12-nt oligomer of cognate single-stranded telomeric DNA ($G_4T_4G_4$) has been solved at high resolution (*28, 29*). The protein complex contains four OB folds, three of which are integral to DNA binding. The Cdc13 DBD exhibits a high degree of structural similarity to each of these OB folds with superpositions of less than 3 Å root mean square deviation (rmsd) over the secondary structural elements of the OB fold (*30*). Notably, the Cdc13 DBD superimposes with a 2.2 Å rmsd to the NH_2-terminal OB fold of the α subunit, the region of TEBP that was used to identify the Pot proteins (Fig. 4). Structure-based sequence alignments revealed no appreciable sequence similarity between the Cdc13 DBD and the *O. nova* OB folds over the region of structural superposition or over the amino acids that make direct DNA contacts. This lack of sequence similarity demonstrates the critical need for structure-based comparisons for assessment of homology among divergent proteins. The close structural relationship observed here suggests that despite its sequence divergence, Cdc13 shares a common ancestor with the *O. nova* TEBP, and therefore with the Pot proteins as well.

The structural similarity between the Cdc13 DBD and other proteins involved in telomere end protection shows that the OB fold is a broadly conserved structural element for binding single-stranded telomeric termini. Functional similarities are also seen in the cellular responses to Cdc13 and Pot1 depletion with regard to chromosome end protection. This combination of structural and functional similarity between Cdc13 and telomere end-binding proteins from other distantly related eukaryotes indicates that mechanisms of telomeric end protection are widely conserved throughout evolution.

References and Notes

1. D. Shore, *Curr. Opin. Genet. Dev.* **11**, 189 (2001).
2. E. H. Blackburn, *Cell* **106**, 661 (2001).
3. T. de Lange, *Science* **292**, 1075 (2001).
4. D. E. Gottschling, V. A. Zakian, *Cell* **47**, 195 (1986).
5. C. M. Price, T. R. Cech, *Genes Dev.* **1**, 783 (1987).
6. P. Baumann, T. R. Cech, *Science* **292**, 1171 (2001).
7. C. I. Nugent, T. R. Hughes, N. F. Lue, V. Lundblad, *Science* **274**, 249 (1996).
8. J.-J. Lin, V. A. Zakian, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13760 (1996).
9. B. Garvik, M. Carson, L. Hartwell, *Mol. Cell. Biol.* **15**, 6128 (1995).
10. T. A. Weinert, L. H. Hartwell, *Genetics* **134**, 63 (1993).
11. C. Booth, E. Griffith, G. Brady, D. Lydall, *Nucleic Acids Res.* **29**, 4414 (2001).
12. S. J. Diede, D. E. Gottschling, *Cell* **99**, 723 (1999).
13. J. Lingner, T. R. Cech, T. R. Hughes, V. Lundblad, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11190 (1997).
14. S. K. Evans, V. Lundblad, *Science* **286**, 117 (1999).
15. E. Pennock, K. Buckley, V. Lundblad, *Cell* **104**, 387 (2001).
16. A. J. Lustig, *Nature Struct. Biol.* **8**, 297 (2001).
17. T. R. Hughes, R. G. Weilbaecher, M. Walterscheid, V. Lundblad, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6457 (2000).
18. E. M. Anderson, W. A. Halsey, D. S. Wuttke, in preparation.
19. Furthermore, as has been previously observed for an NH_2-terminal proteolytic fragment of *S. pombe* Pot1 (*6*) and the α35 subunit of the *O. nova* TEBP (*31*), the Cdc13 DBD binds telomeric DNA more tightly than the full-length protein (*18*).
20. M. S. Singer, D. E. Gottschling, *Science* **266**, 404 (1994).
21. Supplementary Web material is available on *Science* Online at www.sciencemag.org/cgi/content/full/296/5565/145/DC1.
22. S. Dietmann *et al.*, *Nucleic Acids Res.* **29**, 55 (2001).
23. S. Dietmann, L. Holm, *Nature Struct. Biol.* **8**, 953 (2001).
24. A. G. Murzin, *EMBO J.* **12**, 861 (1993).
25. Although our NMR data have allowed for the identification of 11 DNA spin systems, the dearth of NOEs between the spin systems precludes unambiguous assignment of each nucleotide. The low number of internucleotide NOEs is consistent with the DNA adopting a largely extended conformation.
26. Residue numbers listed for the Cdc13 DBD correspond to those of the full-length protein minus 495.
27. The NH_2-terminal OB fold of the *O. nova* α subunit contains a large (18 residue) loop between β2 and β3 that is distant from the site of nucleic acid interaction.
28. M. P. Horvath, V. L. Schweiker, J. M. Bevilacqua, J. A. Ruggles, S. C. Schultz, *Cell* **95**, 963 (1998).
29. M. P. Horvath, S. C. Schultz, *J. Mol. Biol.* **310**, 367 (2001).
30. The *O. nova* α NH_2-terminal domain (35 kD) bound to ssDNA shows no substantial structural differences from the corresponding OB folds in the ternary complex (*32*), validating the use of the ternary complex for structural comparison.
31. G. Fang, J. T. Gray, T. R. Cech, *Genes Dev.* **7**, 870 (1993).
32. S. Classen, J. A. Ruggles, S. C. Schultz, *J. Mol. Biol.* **314**, 1113 (2001).
33. R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **14**, 51 (1996).
34. M. Carson, *J. Appl. Crystallogr.* **24**, 958 (1991).
35. R. M. Mitton-Fry, D. S. Wuttke, *J. Biomol. NMR*, in press.
36. B. T. Farmer *et al.*, *Nature Struct. Biol.* **3**, 995 (1996).
37. G. J. Kleywegt, T. A. Jones, *CCP4/ESF-EACBM Newslett.* **31**, 9 (1994).

38. We thank T. Cech, A. Pardi, and the members of the Wuttke lab for helpful discussions and critical reading of the manuscript; R. Batey, K. Bjorkman, E. Fry, W. Halsey, B. Kelly, C. Mandel, and H. Zhou for technical assistance; and M. Summers and C. Tang for access to an 800-MHz spectrometer. Funded by NIH grants GM59414 (D.S.W.) and GM55867 (V.L), the American Cancer Society (D.S.W.), a University of Colorado Junior Faculty Development Award (D.S.W.), a Beckman Young Investigator Award (D.S.W.), and the U.S. Army Breast Cancer Research Program (E.M.A.). R.M.M.-F. is a Howard Hughes Medical Institute Predoctoral Fellow.

7 December 2001; accepted 15 February 2002

Conserved Structure for Single-Stranded Telomeric DNA Recognition

Rachel M. Mitton-Fry, Emily M. Anderson, Timothy R. Hughes, Victoria Lundblad, and Deborah S. Wuttke

Supplementary Material

Methods. Cdc13 DBD was produced and NMR samples prepared as described in (*1*). In brief, recombinant Cdc13 DBD (residues 497-694, with an NH_2-terminal methionine) was expressed in *E. coli* BL21(DE3) cells. Cell pellets were lysed by French press and protein was purified by ion exchange chromatography over SP-sepharose resin, following incubation of the cell supernatant with 0.1% polyethylenimine to precipitate cellular DNA. Uniformly ^{15}N- or ^{15}N,^{13}C-isotopically labeled protein was prepared by growth in minimal media containing $(^{15}NH_4)_2SO_4$ with or without ^{13}C glucose as the sole nitrogen or carbon sources, respectively. ssDNA (dGTGTGGGTGTG) was purchased from Operon and purified by reversed-phase HPLC. NMR samples contained 0.8-1.5 mM unlabeled, ^{15}N-labeled or ^{13}C, ^{15}N-labeled protein, 0.9-1.7 mM ssDNA, 50 mM imidazole-d_4, pH or pD* 7.0, 150 mM NaCl, 100 mM Na_2SO_4, 0.02% NaN_3, and 2 mM DTT-d_{10} in 10% D_2O/90% H_2O or 100% D_2O. All data were acquired on a 500 or 600 MHz Varian Unity^{INOVA} or 800 MHz Bruker DRX spectrometer at 30 or 35°C. Data were processed using NMRPipe (*2*) and analyzed and assigned using Ansig v3.3 (*3*). Sequential backbone and non-exchangeable aliphatic assignments were obtained from standard heteronuclear NMR experiments and aromatic side chain assignments were obtained from CBHD, CBHE, and NOESY experiments (*4*). Interproton restraints were acquired from a 2D homonuclear NOESY, a 3D ^{15}N-separated NOESY, 3D ^{13}C-separated NOESYs, a 4D ^{15}N, ^{13}C-separated NOESY, and a 4D ^{13}C, ^{13}C-separated NOESY. Mixing times ranged from 100-150 ms. NOE restraints were divided into three distance classes: strong (1.8 to 3.0 Å), medium (1.8 to 3.8 Å), and weak (1.8 to 5.0 Å). The 135 dihedral angle restraints added were derived from TALOS (*5*). 78 hydrogen bond restraints were included on the basis of protection from hydrogen exchange and availability of a potential hydrogen-bonding partner. Intermolecular NOEs were acquired using a 3D ^{13}C-edited Chirp-based select/filter NOESY with a 150 ms mixing time (*6*). Structures were calculated using XPLOR 3.851 (*7*).

References and Notes

1. E. M. Anderson, W. A. Halsey, D. S. Wuttke, in preparation.
2. F. Delaglio *et al.*, *J. Biomol. NMR* **6**, 277 (1995).
3. P. J. Kraulis, *J. Magn. Reson.* **84**, 627 (1989).
4. R. M. Mitton-Fry, D. S. Wuttke, *J. Biomol. NMR*, in press.
5. G. Cornilescu, F. Delaglio, A. Bax, *J. Biomol. NMR* **13**, 289 (1999).
6. C. Zwahlen *et al.*, *J. Am. Chem. Soc.* **119**, 6711 (1997).
7. A. T. Brunger, *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR* (Yale University Press, New Haven, 1993).
8. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).

Supplemental Table 1. Experimental restraints and structural statistics.

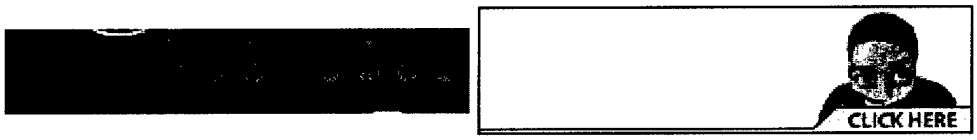
Structural statistics	**Family of ten structures**	
Total restraints (residues 5-191)	2865	
NOEs	2654	
Intraresidue	1210	
Sequential	558	
Medium range ($2 \leq i \leq 4$)	286	
Long range ($i \geq 5$)	600	
Dihedral restraints	135	
Hydrogen bonds	38 x 2	
RMSDs (Å)	**All heavy atoms**	**Backbone atoms**
Residues 7-191	1.74	1.21
OB fold (residues 25-144)	1.62	0.99
OB fold secondary structure*	0.92	0.43
$\beta 2$-$\beta 3$ loop†	2.02	1.18
XPLOR energies (kcal/mol)		
Total energy	735 ± 10	
NOE energy‡	257 ± 12	
Ramachandran analysis (residues 7-191)§		

Residues in most favorable regions	67 %
Residues in favorable regions	23 %
Residues in allowed regions	7 %
Residues in disallowed regions	3 %

* OB fold secondary structure defined as residues 26-38, 44-49, 80-84, 86-99, 122-129, 136-144.
† The family of structures was superimposed over the backbone atoms of residues 25-144 and the RMSDs then calculated for the β2-β3 loop (residues 50-79).
‡ No NOEs were violated by more than 0.5 Å in the final ensemble of structures.
§ Ramachandran analysis performed using Procheck (*8*).

- Similar articles found in: SCIENCE Online
- Abstract of this Article
- Full Text of this Article
- Alert me when: new articles cite this article
- Download to Citation Manager

Volume 296, Number 5565, Issue of 5 Apr 2002, p. 145.



▲ PAGE TOP

Biological Sciences, Biochemistry

Site-directed mutagenesis reveals the thermodynamic requirements for single-stranded DNA recognition by the telomere-binding protein Cdc13

Emily M. Anderson and *Deborah S. Wuttke

Department of Chemistry and Biochemistry

University of Colorado

Boulder, Colorado, 80309-0215

Corresponding author:

Deborah S. Wuttke

Department of Chemistry and Biochemistry

UCB 215

University of Colorado

Boulder, Colorado, 80309-0215

Deborah.Wuttke@colorado.edu

Phone: 303-492-4576/ Fax: 303-492-5894

Manuscript Information:

Text pages: 23 Figures: 4 Tables: 1

Number of Words: 5356 Number of Characters: 47,000

Abbreviations footnote: ss, single-stranded; DBD, DNA-binding domain; OB, oligonucleotide-oligosaccharide; WT, wild type

ABSTRACT

The essential *Saccharomyces cerevisiae* protein Cdc13 functions by binding the conserved single-stranded overhang at the end of telomeres and mediating access of protein complexes involved in both end capping and telomerase activity. The single-stranded DNA-binding domain (ssDBD) of Cdc13 exhibits both high affinity (3 pM) and sequence specificity for the GT-rich sequences present at yeast telomeres. We have used the ssDBD of Cdc13 to understand the sequence-specific recognition of extended single-stranded DNA. The recent structure of the Cdc13 DNA-binding domain revealed that DNA is recognized by a large protein surface containing an OB (oligonucleotide/oligosaccharide-binding) fold augmented by an extended 30-amino acid loop. Contacts to single-stranded DNA occur via a contiguous surface of aromatic, hydrophobic and basic residues. A complete alanine scan of the binding interface has been used to determine the contribution of each contacting sidechain to binding affinity. Substitution of any aromatic or hydrophobic residue at the interface was deleterious to binding (20 to > 700 fold), while tolerance for replacement of basic residues was observed. The important aromatic and hydrophobic contacts are spread throughout the extended interface, indicating that the entire surface is both structurally and thermodynamically required for binding. While all of these contacts are important, several individual substitutions to alanine that abolish binding cluster to one region of the protein surface. This region is vital for recognition of four bases at the 5' end of the DNA and constitutes a "hotspot" of binding affinity suitable for the recognition of the heterogeneous sequences at yeast telomeres.

The recognition of single-stranded nucleic acids plays an important role in many essential cellular processes, including telomere regulation (1), DNA replication and repair (2, 3), transcription (4), translation (5), and RNA processing (6). Single-stranded DNA (ssDNA) is often bound by proteins in a sequence-independent manner, as in the case of *E. coli* single-stranded DNA-binding protein (SSB) (7) or replication protein A (RepA) (8). However, for some processes, such as transcriptional regulation (9) or telomere replication and end protection (1, 10-12), single-stranded DNA recognition is sequence specific. In contrast to proteins that recognize double-stranded DNA or structured RNA, single-stranded DNA-binding proteins recognize an extended nucleic acid conformation that typically involves numerous protein contacts with the accessible bases. Comparatively little is known about the thermodynamic basis for this mode of recognition relative to our understanding of the protein recognition of double-stranded DNA (13-15) and structured RNA (16-18).

The *S. cerevisiae* telomere-binding protein Cdc13 is used in this study to probe the sequence-specific recognition of ssDNA (1). Telomeres contain repetitive tracts of noncoding DNA containing a GT-rich strand with 5'-3' polarity, terminating in a 3' single-stranded overhang (19, 20). Cdc13 specifically recognizes this single-stranded overhang and has at least two genetically distinct roles in the cell. The essential role of Cdc13 is a telomere-capping function. Cdc13 is a component of a telomere end-protection complex that prevents chromosomal degradation and end-to-end fusion (21). Loss of this capping activity results in resection of the C-rich strand of the chromosome end, leading to cell-cycle arrest (22, 23). In addition to capping the telomere, Cdc13 also mediates the recruitment of the replicative enzyme telomerase to the telomere *in vivo* (24,

25). Loss of this activity results in gradual shortening of the telomere and leads to a delayed lethal phenotype. In contrast to both vertebrates and ciliates, budding yeast telomeres are heterogeneous in sequence and can be described with the consensus sequence $G_{2\text{-}3}(TG)_{1\text{-}6}$ (26, 27). Cdc13 binds yeast single-stranded telomeric DNA with high affinity but does not bind closely related telomeric sequences such as *T. thermophila* (T_2G_4) and human (T_2AG_3) (1). Functionally, Cdc13 must discriminate between other single-stranded DNAs and RNAs transiently present in the nucleus. We are investigating the thermodynamic strategy for achieving this specificity.

The minimal single-stranded DNA-binding domain (ssDBD) of Cdc13 exhibits the same specificity for telomeric DNA as the full-length protein (E.M.A., W. Halsey, and D.S.W., unpublished data). The Cdc13 ssDBD uses an OB-fold topology for DNA recognition (28). The oligonucleotide/oligosaccharide-binding (OB) fold superfamily includes several other single-stranded nucleic acid binding proteins such as *E. coli* SSB, Rho, CspA, CspB, and RepA (7, 8, 29-31). Notably, the structure of Cdc13 is similar to OB folds found in the telomere end-binding protein (TEBP) from the ciliate *O. nova* (32). While there is no sequence relationship between Cdc13 and *O. nova* TEBP, the *O. nova* protein exhibits sequence similarity to the telomere-associated Pot1 proteins found in several species, including humans (12), implying that the OB fold is widely used for recognition of single-stranded DNA at telomeres.

In the context of nucleic-acid binding proteins, the OB fold domain typically interacts with a small ligand (2-5 nucleotides) through interactions with the loops on one face of the β–barrel. Extended, single-stranded GT-rich DNA is recognized by Cdc13 across an elongated cleft formed by this face of the OB-fold barrel and an unusually long

loop between strands 2 and 3 (Fig. 1a). Recently, we have assigned all of the contacts between the protein and DNA (28) (Fig. 1). While it is clear that extensive contacts are made to the bases, at this point it is unclear whether extended aromatic stacking interactions occur as was observed in the structure of *O. nova* TEBP, or how much the extended β2-β3 loop contributes to binding (32, 33).

To better understand the molecular determinants of binding in this system and the role of the unusual β2-β3 loop, we have used site-directed alanine mutagenesis to investigate the thermodynamic contributions to ssDNA binding by each of the sidechains identified at the Cdc13/DNA interface. Using this method, contacts that contribute energetically to binding can be distinguished from amino acids that are simply proximal to bound DNA in the structure. This thermodynamic characterization of sequence-specific recognition of extended ssDNA reveals a large contribution to binding from all of the hydrophobic and aromatic residues at the interface. Unexpectedly, even though the substrate dGTGTGGGTGTG (TEL-11) is symmetric in sequence, the 5' end of the binding site contributes disproportionately to the total binding energy.

MATERIALS AND METHODS

Site-directed mutagenesis of the Cdc13 DBD. Site-directed mutagenesis of the Cdc13 ssDBD (aa497-694) was carried out using suitable primers and the QuickChange mutagenesis kit (Stratagene) and was verified by DNA sequencing of the full gene construct.

Expression and purification of recombinant WT and mutant proteins. Proteins were expressed and purified as descibed previously (34). All mutant proteins expressed in

soluble form and were purified in yields comparable to the WT protein (5-12 mg/liter) except the R140A mutant, which was lower-yielding (0.7 mg/liter).

Filter binding assay of equilibrium dissociation constants. Equilibrium binding reactions were performed with ^{32}P-dGTGTGGGTGTG (TEL-11) at concentrations (50 or 200 pM) 20-fold or more below the dissociation constant using serial dilutions of protein in 5 mM HEPES, pH 7.8, 750 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM Na_2EDTA, 2 mM DTT, and 0.1 mg/mL BSA (bovine serum albumin). The reactions were equilibrated on ice for 60 minutes. Filter binding assays were performed using 96-well MultiScreen MAHA N4550 filter plates (mixed cellulose ester membrane). After prewetting the membrane, samples (80 μL) were loaded and incubated for 1 minute before filtering and the wells washed twice with 200 μL of binding buffer lacking BSA and DTT. Filter-bound counts were scanned by Phosphoimager, quantified (Imagequant), and plots were fit to a standard two-state binding model: $y = C1*(x/(x+K_d))+C2$, where y is number of counts of bound DNA, C1 is the maximum plateau of binding, x is the concentration of protein, K_d is the equilibrium dissociation constant, and C2 is the baseline or background counts. Dissociation constants were determined on different days and/or with different protein preparations in triplicate. K_ds determined from three curve fits are reported as average values with standard deviations. The large standard deviations observed for the weak binders reflect the upper limit in sensitivity of the assay due to protein saturation of the filter.

Determination of salt-dependence (KCl) of WT Cdc13 DBD binding. The salt-dependence of DNA-binding by the Cdc13 DBD to dGTGTGGGTGTG was determined by filter binding using a range of KCl concentrations. As a control to determine that

protein binding by the filter was not salt dependent, samples of a range of BSA concentrations were loaded onto filters at each salt concentration stained with Ponceau S. By visual inspection, the filter binding capacity was not affected by salt under the conditions used. The buffer for each K_d measurement was constant as described above except for variation in KCl concentration. Dissociation constants were obtained at 0.6 M KCl, 0.75 M KCl, 0.8 M KCl, 1.0 M KCl, and 1.2 M KCl.

Gel-shift assay of equilibrium binding. The 11mer TEL-11 was 5'-end labeled using 10U of T4 DNA kinase with 5 μM DNA and 150 mCi/ml γ^{32}P-ATP. A 25 μL reaction was incubated at 37°C for 30 minutes, and unincorporated γ^{32}P-ATP was removed using a microspin G25 column (Pharmacia). Equilibrium binding reactions were performed with ^{32}P-11mer at 200 pM and proteins at 200 nM in 5 mM HEPES, pH 7.8, 75 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM Na_2EDTA, 2 mM DTT, and 0.1 mg/mL BSA. The reactions were equilibrated on ice for 60 minutes. 10 μL of each reaction with a small amount of bromophenol blue tracking dye was loaded on a 20 cm x 20 cm x 1.5 mm, 5% acrylamide, nondenaturing gel prerun at a constant 200 V for 45 minutes. The samples were loaded while running, and the gel was run for another 45 minutes to separate the free and bound species. The gel was dried and visualized by Phosphoimager.

RESULTS

Single-Stranded DNA Binding by Cdc13 Exhibits Log-Linear Salt Dependence. The ssDBD of Cdc13 binds single-stranded telomeric DNA, with a K_d of 3 pM at 75 mM KCl (34). The study of affinities in this regime requires extremely low concentrations of DNA probe and extended exposure times for Phosphimager analysis, and is plagued by low signal-to-noise. We investigated conditions under which screening a large number of

mutants would be experimentally tractable. Binding was characterized to the TEL-11 oligonucleotide, a sequence of DNA that is complementary to the yeast telomerase RNA template and is representative of yeast telomeric sequences (35).

Binding conditions were chosen such that the affinity of WT Cdc13 DBD for TEL-11 was 1 nM. Fig. 2a displays representative filter binding data obtained at 750 mM KCl, along with a plot of this data fit to a standard two-state binding model in Fig. 2b. All measurements were conducted under the same binding conditions such that extrapolation to different salt concentrations was not necessary for comparison.

The DNA-binding affinity of the Cdc13 DBD decreased as the concentration of KCl was increased. The logarithmic plot of the binding dissociation constant versus the KCl concentration is linear with a slope of approximately 5 (inset in Fig. 2b). The linearity of the plot suggests that salt does not perturb the binding interaction in deleterious ways such as denaturation of the protein or formation of alternate structures.

Site-directed Alanine Mutagenesis. The DNA-binding interface defined by amino acids that directly contact DNA (28) (E.M.A., R. Mitton-Fry, and D.S.W., unpublished data) was probed by a panel of alanine mutants (Fig. 1). The mutants exhibited a range of affinities to TEL-11 spanning 3 orders of magnitude, and notably none of the alanine point mutants had higher affinity for DNA than WT protein (Table 1).

Eight aromatic to alanine mutations (shown in green in Fig. 1) were made. All of these sidechains were identified as contacting DNA by intermolecular NOEs (28) except Y66 and Y85, which were predicted to contact DNA on the basis of their location in the protein-DNA interface. These mutations had a dramatic effect on DNA binding (Table 1), ranging from a 20-fold reduction (Y70A) to severe loss of binding (> 650-fold).

Mutation at positions F44, Y61, Y63, Y66, and Y131 had an intermediate effect on binding (100-250 fold). These residues, which are thermodynamically important for binding, are spread across the full length (44 Å) of the DNA-binding interface.

Two hydrophobic residues are located at the protein-DNA interface - I83 and I138. Point mutation of these residues (yellow in Fig. 1), results in differential effects on DNA binding. Binding of the I83A mutant is reduced 70-fold, while binding activity is effectively completely lost in the I138A mutant (>700-fold reduction).

As would be expected from the salt dependence of binding, the DNA-binding surface of Cdc13 contains multiple positively charged residues, including K41, K73, K81, and R140. With the prominent exception of R140, mutation of these residues had relatively modest effects on DNA-binding affinity (< 10-fold). The higher salt conditions used in this study might be expected to attenuate the effect of a charge to alanine substitution. However, a gel-shift of these mutants under low-salt conditions ([KCl]=75 mM, Supplemental Fig. 1) qualitatively supports the high-salt data; the charge-to-alanine mutants have WT or near WT affinity in sharp contrast to the aromatic substitutions described above which bind weakly or not at all. Another basic residue, K134, is in the vicinity of the DNA-binding interface but exhibits no direct physical contact with DNA. As a control, we mutated this lysine to alanine. The K134A mutation had a similar effect on binding as the other lysine mutations, supporting the role of the charged residues in providing a positively charged surface that generally favors DNA binding.

Secondary substitutions. The role of the one of the tyrosines (Y61) and the arginine (R140) at the interface were probed in more detail by two conservative mutations. The substitution of Y61 to phenylalanine resulted in a 15-fold decrease in affinity compared

to a 200-fold reduction in Y61A. This significant rescue of binding activity by conservative substitution indicates that an aromatic amino acid is required for DNA binding at this position.

In sharp contrast to the other charged residues, replacement of arginine 140 with alanine had a dramatic effect on binding. An R140K mutant was used to test whether the guanido group of arginine was mediating a specific interaction with DNA. Surprisingly, binding was completely restored in the R140K mutant, thus indicating that at this position only a charged residue is necessary, possibly for a very specific network of hydrogen bonds.

DISCUSSION

To investigate the individual thermodynamic contributions of amino acids located throughout the protein-DNA interface, a complete alanine scan of the DNA-contact residues of the Cdc13 ssDBD was performed. The binding surface identified by direct contacts to the DNA (28) is composed of aromatic, hydrophobic and positively charged residues (Fig. 1). These contact residues are well conserved in Cdc13 homologues from closely related species of yeast (R. Cervantes and V. Lundblad, personal communication). The amino acids at positions 27, 63, 66, 85, and 131 are always tyrosine or phenylalanine. I83 and I138 are isoleucine or valine, and R140 is strictly conserved. The lysines are less well conserved, although the basic character of the interface is preserved. However, from structural and phylogenetic knowledge alone, it is impossible to determine which interactions are critical for single-stranded DNA binding affinity or specificity.

To deconvolute the thermodynamic components underlying ssDNA binding, we first determined the ionic strength dependence of Cdc13 ssDBD binding to the minimal oligonucleotide TEL-11 (inset in Fig. 2b). The attenuation of binding with increasing ionic strength is consistent with the presence of several positively charged protein residues at the protein-DNA interface that interact with the negatively charged phosphate backbone of DNA (28). The electrostatic surface potential of the Cdc13 ssDBD (Fig. 4) exhibits a strong positive potential on the surface of the protein containing the protein-DNA interface. The observed log-linear salt dependence is typical of protein-DNA interactions and has been well characterized with double-stranded DNA-binding proteins (36-39).

Point mutants of the Cdc13 ssDBD across this interface exhibited a vast range (3 orders of magnitude) of binding affinities for TEL-11 (Table 1 and Fig. 3). In general, the charge-to-alanine substitutions had uniformly small effects on binding, with one exception discussed below. These effects are not unique to our high salt conditions, as the relative binding affinities of the mutants at low salt (75 mM KCl) revealed similar trends in binding affinity (Supplementary Fig. 1). The high salt conditions used for binding reduce the magnitude of the charge-charge interactions, which will contribute more free energy to the tighter binding observed at physiological conditions. The charged residues may direct interaction with the phosphate backbone of DNA, as is seen in the OB-fold protein Trbp111, which recognizes structural elements of tRNA using charged residues and some hydrophobic interactions for binding (40). The energetic effects of the charge substitutions may be additive and contributing to an overall electrostatic surface, as the salt-dependence of the K81A mutant protein was attenuated

by close to 1 in the slope of the log-linear plot (data not shown). It remains to be determined if the charged residues at the interface contribute to the specificity of binding.

A significant exception to the generally small energetic effects of charge-to-alanine mutations was observed at arginine 140. Replacement of the arginine with alanine at this position caused a decrease in binding of over 500-fold, which was completely restored by reversion to lysine. Therefore, a positively charged residue is essential at this position, possibly to nucleate an essential recognition element by hydrogen bonding either to protein itself or to the DNA. In the structure of the complex, R140 interacts with the base of T2 and may contribute both to specificity and affinity at this position. Arginine is often present at the positively charged surfaces of proteins that bind nucleic acids. Although there are cases where arginine is critical for the base-specific recognition of nucleic acids, for example in the Tat/Tar complex (41), the U1A protein (42), and in zinc-finger proteins (43), it is unusual for a critical arginine residue to tolerate replacement by lysine.

In contrast to the modest effects observed in the mutation of charged residues, every individual aromatic and hydrophobic substitution had a pronounced effect on binding affinity. This is similar to the deleterious effects seen on ssDNA binding by mutation of conserved aromatics to alanine in the OB domains of RepA (44) and *E coli* SSB (45) where the conserved critical aromatic residues stack with DNA bases. This theme is also observed in RRM (RNA recognition motif) domains, where mutation of a critical phenylalanine involved in recognition by U1A protein destabilizes its interaction with RNA by 5.5 kcal/mol (46). In comparison, an alanine point mutant of the phenylalanine providing the bulk of the recognition energy of AspRS N-terminal domain

is reduced by 2.1 kcal/mol for binding the tRNA anticodon loop (47). Our structural data clearly reveal intimate interactions directly with DNA bases for every aliphatic and aromatic amino acid mutated here (28) (Fig. 1b). The DNA-binding site is bounded by Y27 and Y70, which recognize bases at the 5' and 3' end, respectively. The remaining structurally defined aromatics, F44, Y131, Y63 and Y61, contact the bases of T4, G5, G7 and G8. Based on the rescue of binding activity by Y61F, aromatic character is required for appropriate recognition at this site, consistent with a possible stacking interaction with thymine 8. The two critical and conserved isoleucine residues at the interface interact with neighboring bases; G3 contacts I138 (removal costs 4.0 kcal/mol), while I83 is part of an ensemble of contacts to T4 (removal costs 2.6 kcal/mol).

The effects of these point mutations are not additive. The $\Delta\Delta G$ for each mutant summed for all sites at the interface greatly exceeds the total WT binding energy (Table 1), suggesting that these interactions are not governed by simple shape complementarity or lock-and-key type interactions. In contrast, separable interactions are more typically features of the recognition of a scaffold such as B-form helical DNA (48) and ordered RNA (49). The strong cooperativity observed between protein sidechains in the Cdc13 ssDBD is indicative of their strong physical linkage (50). Upon binding, the flexible, single-stranded DNA adopts a more ordered state to recognize Cdc13, similar to a folding event. This implies that aromatic amino acid/base interactions may provide the large enthalpic contribution required for a loss of DNA conformational entropy. The hydrophobic nature of the Cdc13/ssDNA interface is more characteristic of a protein/protein interface than a protein/dsDNA interface (15), and is similar to proteins which recognize ssRNA, such as RRM domains (51).

The point mutants that have large effects on binding are spread throughout the DNA-binding interface (Fig. 1), confirming that this large interface is required for high-affinity binding of the minimal DNA 11-mer, including the large β2-β3 loop. From comparison of the relative effects of the aromatic and hydrophobic alanine mutants (Fig. 3) they can be grouped into two subtypes. One group exhibits large effects, with four mutants essentially eliminating binding. Interestingly, the mutants with the largest effects cluster on the side of the protein/ssDNA interface that interacts with 5' end of the DNA (Fig. 3b). This region can be considered a "hotspot" of binding energy (52) and suggest a mechanism via which Cdc13 recognizes the heterogeneous sequences present at yeast telomeres. This hotspot contains recognition regions (stands β1, β3, and β5) that are classical for OB fold recognition of a small stretch of nucleic acid (4 nucleotides). Thus, Cdc13 primarily recognizes the GTGT sequence at the 5' end of its target. This region also exhibits the greatest specificity for ssDNA, as it is the least tolerant to base substitution (E.M.A. and D.S.W., unpublished data). This interaction is augmented by the inclusion of the 30 amino acid loop which contains numerous tyrosines (Y61, Y63, Y66, Y70) that add to the total binding energy in recognizing the 3' end, which is also a TGTG sequence. From this model, an inexact number of intervening G bases would be tolerated, consistent with the sequences observed at yeast telomeres (26). It is interesting to note that the target DNA sequence, dGTGTGGGTGTG, is palindromic, so that recognition of the phosphate backbone must contribute to the correct orientation of the ssDNA in the binding site.

Many of the principles of recognition determined here for the binding of Cdc13 to ssDNA are similar to those used by other single-stranded nucleic acid binding proteins.

The structural and thermodynamic importance of direct aromatic interaction supplanted by hydrophobic contacts with exposed bases has been observed in some of the other nucleic-acid-binding members of the OB family, including AspRS (47), LysRS (53), CspB (31, 54, 55), and RepA (8, 44, 56). While it is common to observe some aromatic residues involved in recognition, a total of 7 in Cdc13 is quite unusual, considering that they are all contributing to binding. In contrast, *O. nova* TEBP uses 10 aromatic residues distributed over 3 OB domains to specifically recognize a single-stranded DNA 12-mer (32). The aromatic and hydrophobic nature of the Cdc13 interface differs from what is used by classical dsDNA binding proteins but it is more similar to extended RNA-binding proteins, such as U1A (51) and TRAP (57). In this case, expansion of the classical OB binding motif with a large β2-β3 loop has been suitably adapted for specific interaction with a relatively long, extended ssDNA. This study highlights the importance of aromatic and hydrophobic amino acids for this type of interaction and clarifies how Cdc13 can specifically recognize the heterogeneous G-rich single-stranded sequences found in yeast telomeric DNA.

ACKNOWLEDGEMENTS

We would like to thank Prof. James Goodrich, Prof. Olke Uhlenbeck and David Atkins for careful reading and comments on the manuscript. We thank Rachel Mitton-Fry, Douglas Theobald, and Wayne Halsey for helpful insights, and Rachel Cervantes and Prof. Vicki Lundblad for Cdc13 homolog sequence information. This work was supported by the NIH (GM59414), the American Cancer Society, a University of Colorado Junior Faculty Development Award, and a predoctoral fellowship from the U.S. Army Breast Cancer Research Program (EMA, DAMD17-99-1-9150).

REFERENCES

1. Nugent, C. I., Hughes, T. R., Lue, N. F. & Lundblad, V. (1996) *Science* **274,** 249-252.
2. Lohman, T. M. & Ferrari, M. E. (1994) *Annu. Rev. Biochem.* **63,** 527-570.
3. Wold, M. S. (1997) *Annu. Rev. Biochem.* **66,** 61-92.
4. Swamynathan, S. K., Nambiar, A. & Guntaka, R. V. (1998) *FASEB J.* **12,** 515-522.
5. Deo, R. C., Bonanno, J. B., Sonenberg, N. & Burley, S. K. (1999) *Cell* **98,** 835-845.
6. Varani, G. & Nagai, K. (1998) *Annu. Rev. Biophys. Biomol. Struct.* **27,** 407-445.
7. Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2000) *Nat. Struct. Biol.* **7,** 648-652.
8. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. (1997) *Nature* **385,** 176-181.
9. Braddock, D. T., Baber, J. L., Levens, D. & Clore, G. M. (2002) *EMBO J.* **21,** 3476-3485.
10. Lin, J. J. & Zakian, V. A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 13760-13765.
11. Price, C. M. & Cech, T. R. (1987) *Genes Dev.* **1,** 783-793.
12. Baumann, P. & Cech, T. Ŕ. (2001) *Science* **292,** 1171-1175.
13. Luscombe, N. M., Austin, S. E., Berman, H. M. & Thornton, J. M. (2000) *Genome Biol.* **1,** 1-37.
14. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. (2001) *Nucleic Acids Res.* **29,** 2860-2874.

15. Jones, S., van Heyningen, P., Berman, H. M. & Thornton, J. M. (1999) *J. Mol. Biol.* **287,** 877-896.

16. Draper, D. E. (1999) *J. Mol. Biol.* **293,** 255-270.

17. Moore, P. B. (2001) *Biochemistry* **40,** 3243-3250.

18. Ramakrishnan, V. (2002) *Cell* **108,** 557-572.

19. Shore, D. (2001) *Curr. Opin. Genet. Dev.* **11,** 189-198.

20. Blackburn, E. H. (2001) *Cell* **106,** 661-673.

21. Lustig, A. J. (2001) *Nat. Struct. Biol.* **8,** 297-299.

22. Garvik, B., Carson, M. & Hartwell, L. (1995) *Mol. Cell. Biol.* **15,** 6128-6138.

23. Booth, C., Griffith, E., Brady, G. & Lydall, D. (2001) *Nucleic Acids Res.* **29,** 4414-4422.

24. Evans, S. K. & Lundblad, V. (1999) *Science* **286,** 117-120.

25. Pennock, E., Buckley, K. & Lundblad, V. (2001) *Cell* **104,** 387-396.

26. Forstemann, K., Hoss, M. & Lingner, J. (2000) *Nucleic Acids Res.* **28,** 2690-2694.

27. Forstemann, K. & Lingner, J. (2001) *Mol. Cell. Biol.* **21,** 7277-7286.

28. Mitton-Fry, R. M., Anderson, E. M., Hughes, T. R., Lundblad, V. & Wuttke, D. S. (2002) *Science* **296,** 145-147.

29. Bogden, C. E., Fass, D., Bergman, N., Nichols, M. D. & Berger, J. M. (1999) *Mol. Cell* **3,** 487-493.

30. Schindelin, H., Jiang, W., Inouye, M. & Heinemann, U. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5119-5123.

31. Schindelin, H., Marahiel, M. A. & Heinemann, U. (1993) *Nature* **364,** 164-168.

32. Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A. & Schultz, S. C. (1998) *Cell* **95,** 963-974.

33. Classen, S., Ruggles, J. A. & Schultz, S. C. (2001) *J. Mol. Biol.* **314,** 1113-1125.

34. Anderson, E. M., Halsey, W. A. & Wuttke, D. S. (2002) *Nucleic Acids Res.*, in press.

35. Singer, M. S. & Gottschling, D. E. (1994) *Science* **266,** 404-409.

36. Barkley, M. D., Lewis, P. A. & Sullivan, G. E. (1981) *Biochemistry* **20,** 3842-3851.

37. Misra, V. K., Hecht, J. L., Sharp, K. A., Friedman, R. A. & Honig, B. (1994) *J. Mol. Biol.* **238,** 264-280.

38. Fogolari, F., Elcock, A. H., Esposito, G., Viglino, P., Briggs, J. M. & McCammon, J. A. (1997) *J. Mol. Biol.* **267,** 368-381.

39. Record, M. T., Jr., Zhang, W. & Anderson, C. F. (1998) *Adv. Protein Chem.* **51,** 281-353.

40. Swairjo, M. A., Morales, A. J., Wang, C. C., Ortiz, A. R. & Schimmel, P. (2000) *EMBO J.* **19,** 6287-6298.

41. Aboul-ela, F., Karn, J. & Varani, G. (1995) *J. Mol. Biol.* **253,** 313-332.

42. Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. (1990) *Nature* **348,** 515-520.

43. Wolfe, S. A., Nekludova, L. & Pabo, C. O. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29,** 183-212.

44. Bastin-Shanower, S. A. & Brill, S. J. (2001) *J. Biol. Chem.* **276,** 36446-36453.

45. Ferrari, M. E., Fang, J. & Lohman, T. M. (1997) *Biophys. Chem.* **64,** 235-251.

46. Blakaj, D. M., McConnell, K. J., Beveridge, D. L. & Baranger, A. M. (2001) *J. Am. Chem. Soc.* **123,** 2548-2551.

47. Eriani, G. & Gangloff, J. (1999) *J. Mol. Biol.* **291,** 761-773.

48. Luscombe, N. M. & Thornton, J. M. (2002) *J. Mol. Biol.* **320,** 991-1009.

49. Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W. & Ramakrishnan, V. (1999) *Cell* **97,** 491-502.

50. Wyman, J. & Gill, S. J. (1990) *Binding and linkage : functional chemistry of biological macromolecules* (University Science Books, Mill Valley, Ca).

51. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994) *Nature* **372,** 432-438.

52. Clackson, T. & Wells, J. A. (1995) *Science* **267,** 383-386.

53. Commans, S., Plateau, P., Blanquet, S. & Dardel, F. (1995) *J. Mol. Biol.* **253,** 100-113.

54. Schnuchel, A., Wiltscheck, R., Czisch, M., Herrler, M., Willimsky, G., Graumann, P., Marahiel, M. A. & Holak, T. A. (1993) *Nature* **364,** 169-171.

55. Schroder, K., Graumann, P., Schnuchel, A., Holak, T. A. & Marahiel, M. A. (1995) *Mol. Microbiol.* **16,** 699-708.

56. Brill, S. J. & Bastin-Shanower, S. (1998) *Mol. Cell. Biol.* **18,** 7225-7234.

57. Antson, A. A., Dodson, E. J., Dodson, G., Greaves, R. B., Chen, X. & Gollnick, P. (1999) *Nature* **401,** 235-242.

58. Nicholls, A., Sharp, K. A. & Honig, B. (1991) *Proteins* **11,** 281-296.

Table 1. Equilibrium dissociation constants (K_ds) of WTCdc13 ssDBD and point mutants. K_d values are reported in nM and are given as the average of three separate determinations and their standard deviation. Changes in the free energy of binding ($\Delta\Delta G$) relative to WT are calculated. Note that, because WT Cdc13 ssDBD binds with an affinity of 1 (nM), all subsequent values are relative as well as absolute.

FIGURE LEGENDS

Fig. 1. The Cdc13 ssDBD with amino acids that define the DNA-binding interface. a. The sidechains of the residues chosen for alanine mutagenesis are indicated with aromatic groups are in green, hydrophobic groups are in yellow, and positively charged residues are in blue. b. Schematic of TEL11 (dGTGTGGGTGTG)-protein contacts defined by NMR (28) (E.M.A., R. Mitton-Fry, and D.S.W., unpublished data) in the same orientation and color scheme as a, with phosphate groups shown as gray circles.

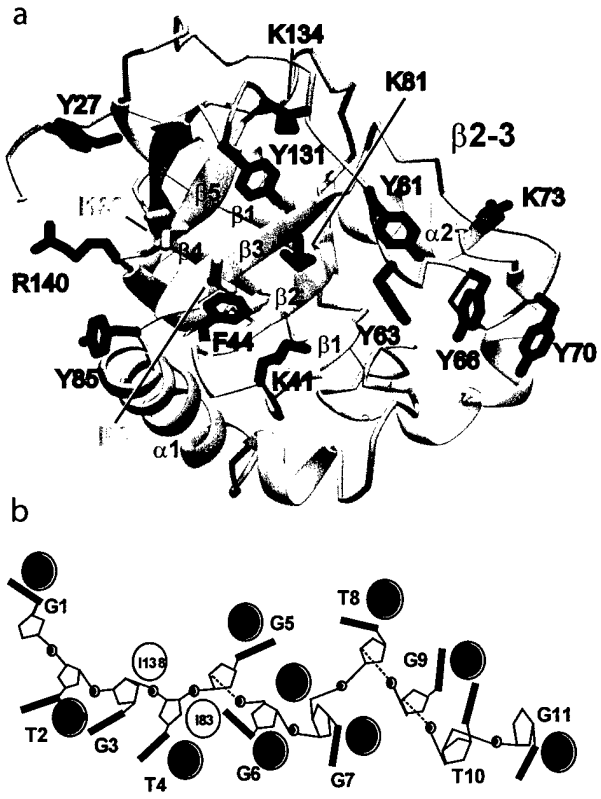
Fig. 2. Filter binding experiments with WT Cdc13 ssDBD. a. Raw filter binding data. The first well contained ^{32}P-labeled DNA probe alone (50 pM), while the remaining wells contained serial dilutions of Cdc13 ssDBD from 100 fM to 2.5 μM. b. A plot of counts bound to the filter (bound DNA) as a function of the concentration of protein. The data were fit to a standard two-state binding model as described in Materials and Methods. The calculated K_d is 1 nM. Inset. A plot of log(K_d, nM) versus log([KCl], M) for WT Cdc13 ssDBD. The data were fit to the line y=0.62 + 5.2x, R=0.97. Each point represents a K_d measurement derived from a separate equilibrium binding experiment.

Fig. 3. Relative binding effects of point mutants. Protein is in the same orientation as Fig. 1. a. The sidechains of the residues mutated to alanine and their relative K_d values for binding dGTGTGGGTGTG are indicated. b. Surface representation of the Cdc13 ssDBD with effects of point mutants scaled according to color. Red represents extreme effects (>500-fold), orange represents large effects (70- to 250-fold), and yellow represents modest effects (3- to 20-fold). The DNA backbone is shown in blue

Fig. 4. a. An electrostatic surface representation of Cdc13 ssDBD. The orientation of the domain is the same as for Figures 1 and 3. Positively charged residues are in blue, and negatively charged residues are in red. The figure was generated using the program GRASP with potential values from -7 to +7 (58).

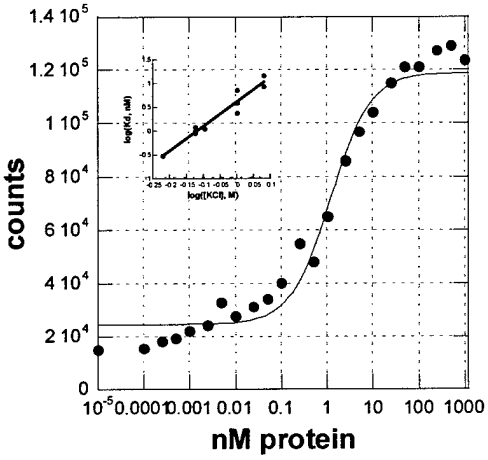
Supplemental Fig. 1. A gel-shift assay in low salt (75 mM KCl, see Materials and Methods) of WT Cdc13 ssDBD and point mutants at concentrations of 50 pM DNA and 200 nM proteins.

mutant	Average Kd (nM)	Std. Deviation	ΔΔG (kcal/mol)
WT	1.0	0.2	-
Y27A	650	210	4.0
F44A	170	68	3.2
Y61A	170	40	3.2
Y61F	14	5	1.6
Y63A	110	36	2.9
Y66A	240	140	3.4
Y70A	23	4.7	1.9
Y85A	680	180	4.0
Y131A	150	63	3.1
I83A	67	11	2.6
I138A	710	260	4.0
K41A	7.9	4.4	1.3
K73A	8.0	2.5	1.3
K81A	2.8	1.2	0.6
K134A	5.7	2.7	1.1
R140A	540	400	3.0
R140K	1.0	0.2	0

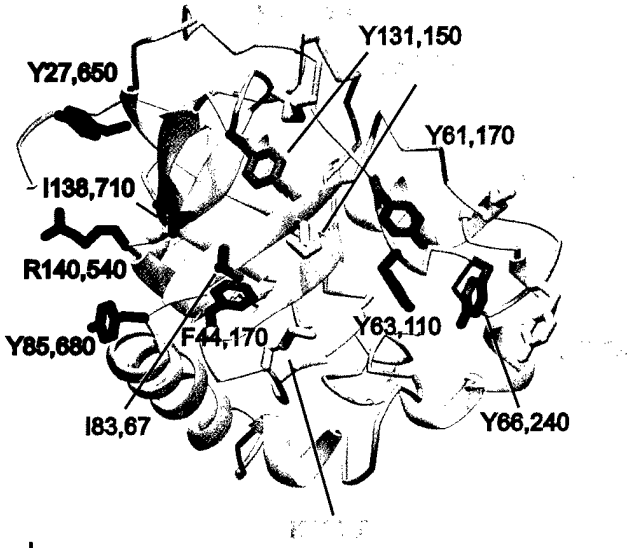

a
K134
K81
Y27
β2-3
Y131
Y61
K73
β1
α2
β4
β3
R140
β2
F44
β1
Y63
Y66
Y70
Y85
K41
α1
b
G1
T8
G5
G9
I138
G11
I83
T2
G3
T4
G6
G7
T10

a

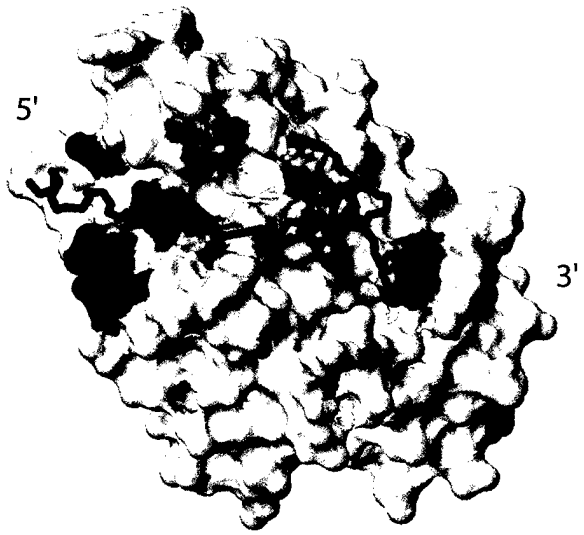
b

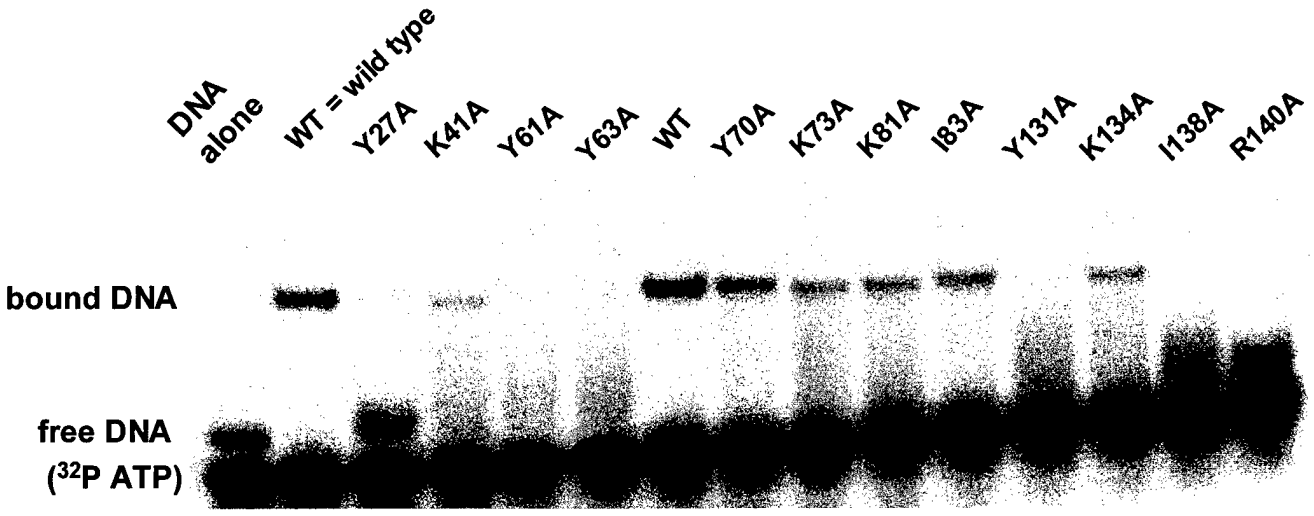

1.4 10^5
1.2 10^5
1 10^5
8 10^4
6 10^4
4 10^4
2 10^4
0
counts
10^-5 0.0001 0.001 0.01 0.1 1 10 100 1000
nM protein

a


Y131,150
Y27,650
Y61,170
I138,710
R140,540
Y85,680
F44,170
Y63,110
I83,67
Y66,240

b


5'
3'


DNA alone
WT = wild type
Y27A
K41A
Y61A
Y63A
WT
Y70A
K73A
K81A
I83A
Y131A
K134A
I138A
R140A
bound DNA
free DNA
(^{32}P ATP)